DOCUMENT RESUME

ED 477 343 EA 032 097

AUTHOR	Bishop, John H.
TITLE	A Prospective Policy Evaluation of the Michigan Merit Award Program.
REPORT NO	PEPG/02-7
PUB DATE	2002-06-10
NOTE	53p.; Paper prepared for the "Taking Account of Accountability: Assessing Politics and Policy" Conference (Cambridge, MA, June 9-11, 2002).
PUB TYPE	Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE	EDRS Price MF01/PC03 Plus Postage.
DESCRIPTORS	*Academic Achievement; Academic Standards; *Accountability; Awards; College Bound Students; Educational Objectives; Foreign Countries; High School Students; High Schools; Incentives; *Merit Scholarships; *Outcomes of Education; Performance; Scores; Secondary Education; State Regulation; *State Standards; Tests
IDENTIFIERS	*Michigan

ABSTRACT

In 1999 Michigan implemented the Michigan Merit Award program, a program to motivate high school students to take their studies more seriously. The program offers 1-year college scholarships to students who meet or exceed state standards on Michigan Educational Assessment Program (MEAP) tests in reading, mathematics, science, and writing. This paper discusses the Michigan program and similar state programs. Section 1 documents the lack of engagement of American high school students and compares the time they devote to schoolwork with the time their foreign counterparts spend on schoolwork. Section 2 assesses the social costs of student disengagement and lack of effort. Section 3 analyzes the structure of the Michigan Merit Award program and shows how it attacks the problem of motivating students to become more engaged in their studies. Section 4 provides evidence on the likely effects of the programs by reviewing studies of other moderate-stakes external-examination systems in other states and several Canadian provinces. Appended are detailed and official descriptions of the Michigan Merit Award program and information on the design and validity of the MEAP, which is used to measure academic achievement in the state. (Contains 31 references.) (WFA)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 477 343

A Prospective Policy Evaluation of the Michigan Merit Award Program.

John H. Bishop

August 2002

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

- This document has been reproduced as received from the person or organization originating it.
- Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

J. Bishop

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

EA 032 097

BEST COPY AVAILABLE

PRELIMINARY DRAFT: DO NOT QUOTE OR CITE WITHOUT AUTHOR'S PERMISSION

A Prospective Policy Evaluation of the Michigan Merit Award Program

By John H. Bishop
Cornell University

PEPG/02-07

jhb5@cornell.edu

Prepared for "Taking Account of Accountability: Assessing Politics and Policy," Kennedy School of Government, Harvard University, June 10 and 11, 2002

EA032091

One of the primary reasons American students learn a good deal less during secondary school than students in other industrialized nations is that they devote less time and intellectual energy to the task.[1] Accountability systems designed to get teachers to try harder and set higher standards will not produce more student learning if [as one high school teacher put it] "students are sitting back in their desks, arms crossed, waiting for their teachers to make them smart (Zoch, 1998, p. 70)."

Learning is not a passive act; it requires the time and active involvement of the learner. In a classroom with 1 teacher and 25 students, there are 25 learning hours spent for every hour of teaching time. Learning takes work and that work is generally not going to be as much fun as hanging out with friends or watching TV. If students cannot be motivated to give up some time socializing or watching TV so that they can learn difficult material and develop high level skills, the time and talents of teachers will be wasted.

An important reason for establishing the Michigan Merit Award program is to motivate secondary school students to take their studies more seriously Other states have chosen to tackle the student motivation problem by requiring students to pass a battery of minimum competency examinations (MCES) before they get a high school diploma. This approach was challenged in Debra P. vs. Turlington, 644F.2d 397 (5th Circuit 1981) and in GI Forum et. al. vs Texas Education Agency. The implementation of Florida's graduation requirement was delayed, but was eventually allowed. The Texas case was decided in the state's favor on January 7, 2000.

Michigan chose not to go down this path largely because it wanted the MEAP HST to reflect more challenging learning goals than would be possible if the MEAP exams were being used to set minimum standards for high school graduation.[2]. It also probably did not want to take the risk that an MCE would lower high school graduation rates and college attendance rates. Instead it took the modest step of putting MEAP HST scores on high school transcripts, something, for example, that Connecticut does with it's CAPT, Ohio does with its 12th grade Tests and New York and North Carolina do with their end-of-course exams.

In 1999 Michigan took the further step of offering a one year $2500 scholarship to students who meet or exceed "Michigan standards" on four MEAP HST tests: Reading, Mathematics, Science and Writing. Students who attend college in Michigan are eligible for the full $2500. Students going to college out of state can receive up to $1500. Starting with the high school

graduating Class of 2005, students have an opportunity to earn up to $500 in addition to the $2,500 they can earn from the Michigan Merit Award, bringing the total possible Michigan Merit Award to $3,000. This award program is based on a student taking all four MEAP tests (Math, Science, Reading and Writing) offered in 7th and 8th grade, and meeting or exceeding state standards on at least two of the four tests.

The Merit Award program is intended:

> **First...to recognize and reward Michigan students who play by the rules, study hard, achieve on their tests and meet high standards. Second, by making postsecondary education more affordable, it encourages students to stay in school and pursue additional education and training after high school. Third, as the "Michigan Merit Award becomes a "household name" in Michigan, even more students will be inspired to raise their performance because they will know the scholarship is available to anyone who is willing to study hard and achieve. Finally, by creating a meaningful incentive for schools to excel and by motivating parents to demand a high quality education for their children, the scholarship program will promote improved school performance in the state.**[3]

This is an ambitious set of objectives for a program whose annual budget is considerably less than 1 percent of total spending on K-12 education in the state of Michigan.[4] Nevertheless, it is well designed for simultaneously achieving all four of these objectives. It has every chance of significantly raising student effort levels, increasing high school completion and college attendance rates, improving the educational climate in most schools and strengthening the resolve of parents and teachers to improve school performance. The key design decision that allows the program to simultaneously serve all four objectives is the decision to base awards on MEAP achievement examinations that reflect the state's recommended curriculum and are graded by the state's teachers. If the awards had been based on a predictive aptitude test like the ACT that is poorly aligned with the state's curriculum, the demand for Kaplan ACT prep courses would have risen but parental pressure for educational excellence would not have been stimulated and school climate would not improve. If awards had been based on high school GPA, objectives 1 and 2 might have been served to some degree, but many students would have responded by choosing unchallenging

courses where A's are easy to get. Most importantly, there would be no incentive for schools to become better and for teachers to set higher standards. To the contrary, pressure on teachers to inflate grades would have intensified.

The paper is organized in four sections. In the first section I document the lack of engagement of American secondary school students and compare the time they devote to schoolwork to the time their overseas counterparts spend on schoolwork. Section 2 assesses the social costs of student disengagement and lack of effort. Students who blow off high school pay a very high price; a much larger price than they imagine when they are in school. They imagine they will be able to go to college regardless of low grades, regardless of low achievement. But, in fact, their chances of completing a degree program are almost zero. They are also unaware that applying themselves in high school helps them get jobs that offer training and promotion opportunities and eventually higher wage rates. Section 3 analyzes the structure of the Merit Award program and shows how it attacks the problem of motivating students to become more engaged in their studies. Section 4 provides evidence on the likely effects of the program by reviewing studies of other moderate stakes external examination systems in other states and in a number of Canadian provinces.

I. The Student Motivation Problem

No matter how you look at it, American secondary schools have a serious student motivation problem. At the completion of his study of American high schools, Theodore Sizer (1984) characterized students as, *"All too often docile, compliant, and without initiative* (p. 54)." John Goodlad (1983) described: *"...a general picture of considerable passivity among students...*(p. 113)." The high school teachers surveyed by Goodlad ranked "lack of student interest" as the most important problem in education.

Time on Task: The low effort levels of American students also evidence themselves in studies of time on task. Classroom observation studies have found that students actively engage in a learning activity for only about half the time they are scheduled to be in school. A study of schools in Chicago found that public schools with high-achieving students averaged about 75 percent of class time for actual instruction; for schools with low achieving students, the average was 51 percent of class time (Frederick, 1977). Overall, Frederick, Walberg and Rasher (1979) estimated 46.5 percent of the potential learning time is lost due to absence, lateness, and inattention.

Studies of time allocation using the reliable time diary method have found that the average number of hours per week in school is 25.2 hours for primary school pupils, 28.7 hours for junior high students and 26.2 hours for senior high students. The comparable numbers for Japan are 38.2 hours for primary school, 46.6 hours for junior high school and 41.5 hours for senior high school (Juster and Stafford 1990). Since studies have found learning to be strongly related to time on task (Wiley 1986; Walberg 1992), these large differentials in time committed to learning are an important reason for the lag of American students behind Japanese students in math and science.

Homework: Harris Cooper's (1989) meta-analysis of randomized experimental studies found that students assigned homework scored about one-half a standard deviation higher on post tests than students not receiving homework assignments. The impact of homework on the rate at which middle school students learn was also significant, though somewhat smaller. There was no evidence of diminishing returns as the amount of homework assigned increased. Nonexperimental studies indicate that the relationship between homework and learning is linear.

Nevertheless, homework is not even assigned in some classes. Arthur Powell describes one school he visited:

> *Students were given class time to read The Scarlet Letter, The Red Badge of Courage, Huckleberry Finn, and The Great Gatsby because many would not read the books if they were assigned as homework. Parents had complained that such homework was excessive. Pressure from them might even bring the teaching of the books to a halt....[As one teacher put it] "If you can't get them to read at home, you do the next best thing. It has to be done....I'm trying to be optimistic and say we're building up their expectations in school."*(Powell, Farrar and Cohen 1985, p.81)

In the High School and Beyond Survey, students reported spending an average of only 3.5 hours per week on homework (National Opinion Research Corporation 1982). Time diaries yielded similar estimates for the early 1980s: 3.2 hours for junior high school and 3.8 hours for senior high school. Time diaries for Japanese students reveal that they spent 16.2 hours per week studying outside of school in junior high school and 19 hours a week studying in senior high school (Juster and Stafford 1992).

Homework assignments have increased since the early 1980s but hours spent doing homework remain low. In a 1991 survey, 29 percent of American 13 year olds said they were doing two or more hours of homework daily. The proportion doing more than two hours of homework was equally low in Canada and Portugal and even lower in Scotland and Switzerland. In most counties, however, the proportion was higher: 79 percent in Northern Italy, 63-64 percent in Ireland and Spain, 50-58 percent in Israel, Hungary, France, Jordan and the former Soviet Union and 41-44 percent in Brazil, Korea, Taiwan and China (NCES 1992b Table 387).

A remarkably large number of students do not do the homework they are assigned. In the Educational Excellence Alliance's (EEA) survey of 21,535 students in Connecticut, Massachusetts, New Jersey and Pennsylvania, only 55 percent said they did all their homework, 29 percent said they did most of their homework and 16 percent said they did none or only some of their homework. When we analyzed who has a high GPA, the single best predictor was the share of homework done, not race, parents education or self-reported ability.

Other Uses of Time: When homework is added to engaged time at school, the total time devoted to study, instruction, and practice in the U.S. is only 18-22 hours per week -- between 16 and 20 percent of the student's waking hours during the school year. By way of comparison, the typical high school senior spent nearly 10 hours per week in a part-time job (NORC 1982) and 19.6 hours per week watching television. Thus, TV occupies as much time as learning.

While some students are overscheduled and find it difficult to fit homework into their busy schedule, most have lots of free time. In the EEA survey 58 percent of students said they spend

two or more hours per day watching TV. Fifty-two percent said they spend two or more hours a day talking with friends and hanging out.

Numerous studies conducted in a variety of countries have found that time spent watching TV is negatively correlated with student performance in school (IAEP 1992). In Table 1 we can see that secondary school students in other industrialized nations watch much less television: 55 percent less in Finland, 70 percent less in Norway and 44 percent less in Canada. Note that in other countries high school students watch less TV than adults; in the United States they watch more. Reading takes up 6 hours of a Finnish student's non-school time per week, 4.8 hours of Swiss and Austrian students time but only 1.4 hours of an American students time.

Peer Pressure against Studying: Probably the most important reason for lack of student engagement is a peer culture that is often hostile to studiousness and public displays of enthusiasm for academic learning.

Steinberg, Brown and Dornbusch's recent study of nine high schools in California and Wisconsin concluded that:

> ...less than 5 percent of all students are members of a high-achieving crowd that defines itself mainly on the basis of academic excellence... Of all the crowds the 'brains' were the least happy with who they are--nearly half wished they were in a different crowd.[5]

Why are the studious called ***suck ups***, ***dorks*** and ***nerds*** or accused of "***acting white***"? Why are students who disrupt the class or try to get the class off track, not sanctioned by their classmates. In part, it is because many teachers grade on a curve and this means trying hard to do well in a class is making it more difficult for others to get top grades. When exams are graded on a curve or college admissions are based on rank in class, joint welfare is maximized if no one puts in extra effort. In the repeated game that results, side payments--friendship and respect--and punishments—ridicule, harassment and ostracism--enforce the cooperative "don't study much, hang out instead" solution. If, by contrast, students were evaluated relative to an outside standard, they would no longer have a personal interest in getting teachers off track or persuading each other to refrain from studying. Peer pressure demeaning studiousness might diminish.

Student Preference for Easy Courses: Although research has shown that learning gains are substantially larger when students take honors and AP courses,[6] only a minority enroll in these courses. In many schools guidance counselors allow only a select few into these courses. Many students prefer easy courses. In a 1987 survey, 62 percent of 10th graders agreed with the

statement, ***"I don't like to do any more school work than I have to."*** [7] Parents often agree with their child. As one guidance counselor described:

> ***A lot of... parents were in a 'feel good' mode."...If they [the students] felt it was too tough, they would back off. I had to hold people in classes, hold the parents back. [I would say] "Let the kid get C's. It's OK. Then they'll get C+'s and then B's." [But they would demand,] "No! I want my kid out of that class!"*** [8]

Rigorous courses are avoided because the rewards for the extra work are small for most students. While selective colleges evaluate grades in the light of course demands, many colleges have, historically, not factored the rigor of high school courses into their admissions decisions. Trying to counteract this problem, college admissions officers have been telling students that they are expected to take the most rigorous courses offered by their school. This effort has met with some success. More students are taking chemistry and physics and advanced mathematics. The bulk of students, however, do not aspire to attend selective colleges and for them, avoiding rigorous courses and demanding teachers is a reasonable strategy.

Pressure on Teachers to Lower Standards: Whether they admit it or not, most teachers explicitly or implicitly grade on a curve. Students are being evaluated relative to the other members of the class, not against an external standard. When a teacher is unsuccessful at teaching a topic, he can leave it off the exam. When students fail to do a good job on an assignment, the teacher can adjust the standard they apply in grading the work. Normally, the struggle over expectations plays out in the privacy of the classroom. Sizer's description of Ms. Shiffe's biology class, illustrates what sometimes happens:

> She wanted the students to know these names. They did not want to know them and were not going to learn them. Apparently no outside threat--flunking, for example--affected the students. Shiffe did her thing, the students chattered on, even in the presence of a visitor....Their common front of uninterest probably made examinations moot. Shiffe could not flunk them all, and, if their performance was uniformly shoddy, she would have to pass them all. Her desperation was as obvious as the students' cruelty toward her. (Sizer, 1984 p. 157-158)

Some teachers are able, through the force of their personalities, to induce their students to undertake tough learning tasks. But for all too many, academic demands are compromised because the bulk of the class sees no need to accept them as reasonable and legitimate.

When teachers try to set high standards, they often get pressured by parents and administrators to go easy. Thirty percent of American teachers say they "feel pressure to give

higher grades than students' work deserves." Thirty percent also feel pressured "to reduce the difficulty and amount of work you assign."[9] Sometimes they get fired (Ann Bradley, Education Week. September 19 1993, p. 1, 19, 20)

II. The Social Costs Of Student Disengagement And Lack Of Effort

Who suffers when students fail to devote sufficient time and energy to learning in high school? Who suffers when teachers set undemanding standards? Not corporate America, they can respond to shortages of skilled workers by moving critical functions abroad and simplifying the jobs that stay in the U.S. Profits need not decline. Not the teachers. They keep their job and avoid being hassled by parents and administrators for handing out poor grades and failing students. It is the students who lose. They lose in two ways.

First, their college aspirations end up not being fulfilled. Just about everybody wants to go to college—even those with poor grades and low test scores. Completing a college program, however, depends on the quality of the student's preparation in high school. For high school sophomores who tested in the top quartile in 1980, 62 percent actually got a bachelors degree in the next 12 years and another 7.2 percent got an associates degree. What about students in the bottom quartile of the test score distribution? Seventy five percent of them said, when they were high school sophomores, that they intended to go to college. But, twelve years later only 3.3 percent of them had actually obtained a bachelors degree and only 4.1 percent had gotten an Associates degree. Other student background characteristics—parent's education, race, socio-economic status also influence the probability of going to and completing college but none hased as powerful an effect on actual outcomes.[10] Many students appear to believe that they do not need to apply themselves in high school to achieve their goal of going to and completing college. They know that a local college will admit them even if they don't know how to spell or write a coherent paragraph. What they do not realize is that if they have not developed these and other basic skills in high school, actually completing a degree program is going to be extremely difficult.[11]

Low achievers will also pay a price by having to work in low wage jobs offering little job security and few chances for advancement. We seldom measure the actual literacy levels of adults but when we do we find that literacy has at least as big an effect on earnings and unemployment as years of schooling. Table 2 presents evidence for this assertion from the National Adult Literacy Survey. Adults in the bottom prose literacy group earn one-third as much as those in the top literacy

group and were 6.5 times more likely to be unemployed. High school dropouts, by contrast, earned 43 percent of what college graduates earn and were 2.6 times more likely to be unemployed.[12]

Altonji and Pierret's study of how scores on the Armed Forces Qualification Test (AFQT) taken while a teenager effect subsequent labor market success provides estimates of the magnitude of the effects of literacy and basic skills in the late 1980s and early 1990s. They are presented in Figure 1. Controlling for a contemporaneous measure of completed schooling, they found that a one standard deviation (4-5 grade level equivalent) higher AFQT score was associated with only a 2.8 percent increase in wage rates the first year out of school but a **16 percent** increase 11 years later.[13] By contrast, the percentage impact of a year of schooling decreased with time out of school from 9.2 percent for those out just one year to 3 percent for those out for 12 years.

Literacy's effect on wages is initially small because employers seldom know which job applicants have the literacy skills they seek. Over time, however, employers learn which employees are the most competent by observing job performance. Those judged most competent are more likely to get further training, promotions and good recommendations when they move on. Poor performers are encouraged to leave. Since academic achievement in high school is correlated with job performance,[14] the sorting process results in basic skills assessed during high school having a much larger effect on the labor market success of 30 year olds than of 19 year olds.[15]

The long delays before the benefits of academic achievement in high school start accruing send students the wrong signal. Teenagers know that college educated adults have good jobs and live in large attractive houses. That's why so many want to go to college. They do not know whether the successful adults they see in their community took rigorous courses and studied hard in high school. As we saw above they will observe almost no relationship between academic achievement of their older siblings/friends and the quality of their jobs. So it would be reasonable for youngsters to conclude that while credentials are rewarded by employers, learning is not. If that is the conclusion they draw, the best strategy for the bulk of students is to study just hard enough to get the diploma and be admitted to college, but no harder.

III. Motivating Students To Pay Attention In Class And Study Harder?

How can incentives for classroom engagement and hard study be increased? Lets begin by examining what student say motivates them to work hard in school. In 1998/99 the Educational Excellence Alliance (EEA) surveyed 35,000 students in 135 high schools in New York, New Jersey, Massachusetts, Connecticut and Pennsylvania. Students were asked "**When you work really hard in school, which of the following reasons are most important for you?**" The most frequently cited reasons were extrinsic and future oriented.

- "I need the grades to get into college"..... 79 %
- "Help me get a better job"..... 58 %

Parents came in second:

- "To please or impress my parents"..... 55%
- "My parents put pressure on me"..... 44%

Intrinsic motivation placed third.

- "The subject is interesting"..... 42 %

Teachers came in fourth:

- "My teachers encourage me to work hard"... 31 %
- "The teacher demands it"..... 22 %
- "To please or impress my teacher"..... 22 %

Multiple regression analysis of the EEA data confirmed the finding that prospects of going to college were the single most important reasons for working in high school. Holding other characteristics of the student body constant, schools with large numbers of students citing "need the grades to get into college" as their reason for working hard tended to have higher levels of classroom engagement and fewer students not doing their homework.

Some have proposed to strengthen incentives to study in high school by raising the minimum academic standard students must reach before they will be admitted to any post-secondary institution. This would be unwise for three reasons. Most people feel that society should offer everyone, no matter their age or how many mistakes they have made in the past, the opportunity to go back to school and try to make something better out of the rest of their life. The adolescent culture of high schools makes them alien territory for adults. Only colleges with open door admissions policies can serve this 2nd chance, 3rd chance function. Secondly, Michigan has set

a goal of expanding participation in post-secondary education. Ending open-door admissions policies might prevent that objective from being realized. Finally, denying admission to all colleges [not just one particular college] is clearly a high stakes decision. One would not want to base such an important decision solely on test scores from a single battery of tests.

Michigan has chosen a much wiser course. It's Merit Award program is well designed to simultaneously induce parents and teachers to set higher standards, induce students to study harder and increase college attendance rates. It has a number of positive features.

1. Conditioning awards on achievement makes absolutely transparent what students and parents must do to seize the opportunity. Students are being urged to study harder and to sign up for more demanding courses. The extra learning this produces benefits the student regardless of whether she ends up getting a merit award. By contrast, need-based financial aid programs often send no signal, a murky signal or the wrong signal and stimulate undesirable behavior. The rules for determining eligibility for need-based financial aid are highly complex and vary from institution to institution. Many low and moderate income parents are not aware that generous need-based financial aid will be forthcoming if their child is admitted to University of Michigan or Michigan State University, so they do not urge their children to set their sights high and to build the kind of academic record that would get them into the state's flagship institutions. At the other end of the spectrum are the growing number of savvy parents who arrange their finances to maximize their eligibility for financial aid. Here are some of the strategies recommended by the financial aid guide books:

   < Do not create an education trust fund in your child's name. Financial aid formulas tax such assets at an extremely high rate.

   < Put as many as possible of your own assets into 401k plans and IRAs. These are not counted as assets in financial aid formulas.

   < In the year before your child enters college, Minimize your adjusted gross income on your federal tax return by having a Schedule C business with lots of deductible expenses.

2. Scholarship eligibility is open ended. The award goes to every student who meets or exceeds the absolute standard. Everyone in the school has the potential of getting the scholarship; not just the best student in French or Music or the students who rank in the top 10 percent of the graduating class. These other kinds of merit scholarships have the dysfunctional effect of pitting classmates against each other. Students who win these traditional merit scholarships are honored by their parents, but their classmates see them as nerds, suck ups or "Oreos."

That is why many schools stopped awarding these honors at compulsory daytime school assemblies. There were too many incidents of catcalls mixed with unenthusiastic applause. The Merit Award, by contrast, helps to reduce anti-nerd peer pressure. Students who joke around in class or try to get the teacher off track will no longer be honored and rewarded by peers because their disruptions make it harder for the rest of the class to get the $2500 award.

3. Basing the Merit Award on an external assessment brings the educational goals of students, parents and teachers into alignment. Prior to the Merit Award program students and parents benefited little from administrative decisions opting for higher standards, more qualified teachers or a heavier student workload. The immediate consequences of such decisions--higher taxes, more homework, having to repeat courses, lower GPA's, complaining parents, a greater risk of being denied a diploma--were negative. As a result, parents pressured teachers to be easy graders and were reluctant to vote higher tax levies so more highly qualified high school teachers could be recruited. The Merit award program will make parents stronger advocates of higher standards and better teaching.

4. The Merit Award standard was set at a level that is achievable by almost all students. The cut point is in the fat middle part of the distribution of student achievement, so incentive effects are maximized. Few will feel the HST tests are so difficult, they have no chance of being recognized for meeting Michigan standards.

5. The special long term financing of the program means that parents of 9 year olds can be confident it will be there for their child when she finishes high school. This maximizes incentive effects because confidence in the future availability of the Merit Award improves student attitudes and effort throughout their school career not just during the junior and senior year of high school. This is one of the reasons why "I Have a Dream" programs often have such salutary effects on student motivation and success.[16] Studies of the impacts of need-based financial aid have found strong effects on which college students attend, but they have not been able to establish that it has large effects on the overall college attendance rates. One reason for this second finding may be that key decisions are made in middle school about courses taken and how hard to try and middle school students from low income families are unaware that they will be eligible for generous financial aid if they build a solid academic record.

6. The centralized grading of the extended answer portions of MEAP exams by Michigan teachers is a very positive feature of the program. Having to agree on what constituted excellent, good, poor, and failing responses to essay questions or open-ended math problems results in a sharing of perspectives and teaching tips that the teachers find very helpful. In May 1996 I interviewed a number of teachers union activists about the examination system in the Canadian province of Alberta. They universally reported that serving on grading committees was "...a wonderful professional development activity (Bob, 1996)." The opportunity to grade the writing exam and the extended answer parts of other exams should be rotated among teachers so that most teachers get this very valuable professional development experience.

7. The scholarship is modest in size and lasts for only one year. Consequently the selection of scholarship winners is a low or moderate stakes decision not a high stakes decision.[17] Because the stakes are moderate not high, the APA's recommendation that decisions be made on the basis of multiple indicators does not apply to the award of merit awards on the basis of MEAP test scores. [18]

8. No one is made worse off. In fact, those who do not meet Michigan standards and do not get a Merit award will find it easier to get conventional need-based aid. Michigan colleges will tend to be redirect their budgets for student assistance towards those not eligible for Merit awards.

9. The Merit Award Program is a small part of an integrated and balanced system of financing higher education and assisting students to attend college. Many of the other components of this funding system target their funds on disadvantaged and minority students. Families with incomes below $100,000 are eligible for a federal tax credit of up to $1500 for each student going to college. This probably yielded Michigan families about $366,000,000 in tax credits last year.[19] In addition to the tax credits, federal student aid programs provided Michigan undergraduates $235,206,000 in need based grants and interest subsidies in fiscal 1997 (NCES, 1998, Table 365, p. 416). Institutions of higher education in Michigan awarded $466,289,000 in scholarships and grants in fiscal 1996 much of which was need-based and went to undergraduates. The Merit Award program adds about $100,000,000 annually to the student aid pot, less than one-tenth of the total. In

addition, state and local government appropriated $1,927,812,000 to support higher education institutions in fiscal 1996.[20] Almost all of these funds support the instructional function of these institutions and directly benefited students. The state funding was roughly $4,727 per student. Without these state funds, tuition would have doubled or tripled, pricing many low and moderate income students out of college. College students also benefit from a host of other state and federal subsidies: the deductibility of gifts to higher education and the tax exempt status of land and buildings and endowment income. Thus the Merit Award program is just a tiny piece--3.4 percent--of total public subsidies of higher education in the state of Michigan. It's the merit piece of an overall higher education funding plan that devotes more than six times as much money to need-based student financial aid.

10. The Merit Award program improves the effectiveness of the state's efforts to hold high schools accountable for student achievement. When a test is not part of a course grade or important to the student in some other way, many high school students fail to put much effort into answering all the questions correctly and completely.[21] Prior to the MEAP Merit Award Michigan students had no reason to try hard on MEAP HST tests, the primary indicator of student achievement in the state's high school accountability system. Many students were boycotting the test. School ratings thus reflected, in part, a high school's success in getting students to try hard on HST tests. This reduced the validity of high school tests as measures of true student achievement and tended to make their use in accountability systems problematic. The MEAP Merit Award has given students an incentive to do the best they can on the HST tests and this has improved the fairness and validity of the state's school accountability system.

11. The MEAP assessment is a much better exam than the tests that most teachers develop for themselves and use to grade their students. It is the product of an extensive consultative process. Input was obtained from hundreds of teachers—teachers who were highly respected by their colleagues. All items are pre-tested and reviewed for ambiguity and bias by trained testing professionals. (see Appendix B for a complete description of the development process) The tests that teachers develop for themselves are, by contrast, generally of very low quality. Fleming and Chambers (1983) study of tests developed by high school teachers found that "over all grades, 80% of the items on teachers' tests were constructed to tap the

lowest of [Bloom's] taxonomic categories, knowledge (of terms, facts or principles)"(Thomas 1991, p. 14). Rowher and Thomas (1987) found that only 18 percent of history test items developed by junior high teachers and 14 percent items developed by senior high teachers required the integration of ideas. College instructors, by contrast, required such integration in 99 percent of their test items. Most secondary school teachers test for low level competencies because that is what they teach. I have reviewed the released items for the MEAP High School Tests in Mathematics, Reading, Science, and Writing and they appear to me to be pushing instruction in right direction.

12. The Merit Award program tends to redirect student energy away from preparing for high stakes multiple choice tests like the SAT-I and the ACT and toward the learning the curriculum that Michigan has developed for its students. This is a good thing because the ACT and the SAT-1 are not comprehensive measures of learning during high school.[22] The energy that students devote to cracking the ACT would be better spent reading widely and learning to write coherently, to think scientifically, to analyze and appreciate great literature and to converse in a foreign language. These are the true objectives of a high school education. The high stakes attached to the ACT and the SAT-1, however, tend to direct student energy away from developing these important skills and weakens the ability of teachers to set high standards themselves. The MEAP High School Tests are much better assessments of Michigan's curriculum objectives than the ACT. The MEAP tests have been developed with great care and are far superior to the ACT test upon which carries such high stakes for Michigan students.

The Merit Award program is well designed to achieve its objectives of stimulating greater student effort and raising academic standards. What do the experiences of other states and nearby Canadian provinces tell us about its likelihood of success. We turn now to a review of that evidence.

IV. The Effects of Moderate Stakes Curriculum-Based External Exit Exams on Student Achievement and High School Climate

How has the Merit Award program changed the incentives faced by Michigan students? Prior to the Merit Award program, the measures of student competence that were rewarded were ACT test scores and grade point averages. MEAP HSTs were no-stakes exams and many students were blowing them off. What has changed? First, the rewards for learning increased. Second, the Merit Awards changed how student achievement was defined and rewarded. ACTs scores and GPAs still matter but now state-developed curriculum-based external assessments of achievement matter as well. By this step Michigan created a low/moderate stakes curriculum-based external exit exam system. What's a curriculum-based external exit exam system (CBEEES)? It:

1. **Produces signals of student accomplishment that have real consequences for the student. While some stakes are essential, high stakes may not be necessary.** Analyses of Canadian and US data summarized below suggest that moderate stakes may be sufficient to produce substantial increases in learning.
2. **Defines achievement relative to an external standard, not relative to other students in the classroom or the school.** Fair comparisons of achievement across schools and across students at different schools are now possible. Costrell's (1994) analysis of the optimal setting of educational standards concluded that more centralized standard setting (state or national achievement exams) results in higher standards, higher achievement and higher social welfare than decentralized standard setting (ie. teacher grading or schools graduation requirements).
3. **Is organized by discipline and keyed to the content of specific course sequences.** This focuses responsibility for preparing the student for particular exams on one (or a small group of) teacher/s. Alignment between instruction and assessment is maximized and accountability is enhanced.
4. **Signals multiple levels of achievement in the subject.** If only a pass-fail signal is generated by an exam, the standard will have to be set low enough to allow almost everyone to pass and this will not stimulate the great bulk of students to greater effort (Kang 1985; Costrell 1994).
5. **Sponsored by and developed to the specifications of the department that funds and regulates elementary and secondary education in the state.** Curriculum reform is

facilitated because coordinated changes in instruction and exams are feasible. Tests established and mandated by other organizations serve the interests of other masters. America's premier high stakes exams--the SAT-I and the ACT—serve the needs of colleges to sort students by aptitude, **not** the needs of schools to reward students who have learned what high schools are trying to teach.

6. **Covers all or almost all secondary school students.**

7. **Assess a major portion of what students studying a subject are expected to know or be able to do.** Studying to prepare for an exam (whether set by one's own teacher or by a state department of education) should result in the student learning important material and developing valued skills. Some MCEs, CBEEES and teacher exams do a better job of achieving this goal than others. External exams, however, cannot assess every instructional objective. Teachers should be responsible for evaluating dimensions of performance that cannot be reliably assessed by external means or that local leaders want to add to the learning objectives specified by the state department of education.

High stakes curriculum based external exam systems are found throughout East Asia and in much of Europe—England, Scotland, Ireland, France, Italy, Denmark, Finland, Hungary, Poland, Russia, the Czech Republic and the Slovak Republic. Careful empirical analysis of data from the 40 nation Third International Mathematics and Science Study (TIMSS) has found that teaching is more rigorous and students learn more in nations with CBEEES.[23] Analysis of data from TIMSS found that students from countries with CBEEE systems outperform students from other countries at a comparable level of economic development by **1.3** U.S. grade level equivalents in science and by **1.0** U.S. grade level equivalent in mathematics. A similar analysis of International Assessment of Educational Progress data on achievement in 1991 of 13 year olds in 15 nations found that students from countries with CBEEES outperformed their counterparts in countries without CBEEES by about 2 U.S. grade level equivalents in math and about two-thirds of a US grade level equivalent in science and geography. Analysis of data from the International Association for the Evaluation of Educational Achievement's study of reading literacy of 14 year olds in 24 countries found that students in countries with CBEEES were about **1.0** U.S. grade level equivalent ahead of students in nations at comparable levels of development that lacked a CBEEES.[24]

In some of these nations the stakes attached to exam results are extremely high. It is quite legitimate to question how relevant these findings are for evaluating the likely effects of low and moderate stakes CBEEES systems like the one in Michigan. While most nations with CBEEESs have gone the high stakes route, some have not—Canada and the Netherlands. We will look at Canada. In addition, two American states—New York and North Carolina--have had moderate stakes CBEEES for many years and Connecticut has had a low stakes CBEEES, the CAPT—since 1994.

Evidence from Canada: In 1990-91, the year the IAEP data being analyzed was being collected, Alberta, British Columbia, Newfoundland, Quebec and Francophone New Brunswick had curriculum-based provincial examinations in English during junior year and French, mathematics, biology, chemistry, and physics during the senior year of high school. The other provinces did not have curriculum-based provincial external exit examinations. The exams were developed by teachers selected by the Ministry of Education and graded by teachers in centralized locations. Exam scores accounted for 50 percent of that year's final grade in Alberta, Newfoundland and Quebec and 40 percent in British Columbia. While exam results appeared on transcripts, college admissions decisions were based almost entirely on high school grades and were generally made before the senior year exams were graded. The study found that controlling for the siz and structure of the school and social background of it's students, schools in provinces with CBEEES taught their students a statistically significant one-half of a U.S grade level equivalent more math and science than comparable schools in provinces without CBEEES.[25]

The impacts of CBEEES on school policies and instructional practices were also studied. CBEEES were not associated with higher teacher-pupil ratios or greater spending on K-12 education. They were, however, associated with higher teacher salaries, a greater likelihood of having teachers specialize in teaching one subject in middle school and a greater likelihood of hiring teachers who have majored in the subject they will teach. Schools in CBEEES provinces devoted more hours to math and science instruction and built and equipped better science labs. The number of computers and library books per student were unaffected by CBEEES.[26]

Fears that CBEEES would cause the quality of instruction to deteriorate appear to be unfounded. Students in CBEEES jurisdictions were less likely to say that memorization is the way to learn the subject and more likely to do experiments in science class. Apparently, teachers subject to the subtle pressure of an external exam four years in the future adopted strategies that were

conventionally viewed as "best practice," not strategies designed to maximize scores on multiple choice tests. Quizzes and tests were more common, but in other respects a variety of indicators of pedagogy were no different in CBEEES jurisdictions. They were not less likely to like the subject and they were more likely to agree with the statement that science is useful in every day life. Students also talked with their parents more about school work and reported their parents had more positive attitudes about the subject.

New York and North Carolina's Moderate Stakes CBEEES: Begun in the 1860s, New York State's curriculum-based Regents Examination System is the oldest American example of end–of-course examinations (EOCE). A college bound student taking a full schedule of Regents courses would typically take Regents exams in mathematics and earth science at the end of 9th grade; mathematics, biology and global studies exams at the end of 10th grade; mathematics, chemistry, American history, English and foreign language exams at the end of 11th grade and a physics exam at the end of 12th grade. For students the stakes attached to Regents exams were pretty modest. Each district decided whether Regents exam grades were to be a part of the course grade and how much weight to assign to them. While almost all districts counted Regents exam results as a final exam grade (teachers or departments sometimes gave their own final as well) so when grades on finals were averaged in with quarterly marking period grades, Regents exam scores seldom accounted for more than a quarter of the student's final grade in a course. Eligibility for a "Regents" as opposed to a local diploma depended on passing the Regents exams, but the benefits of getting a "Regents" diploma have declined and have been small for the last two decades. During the 1950s and 60s Regents exam scores were used to select winners of Regents scholarships. Regents exam grades also appeared on high school transcripts, but in recent years college admissions decisions depended primarily on grades and SAT scores, not Regents exam scores or Regents diplomas.[27]

North Carolina introduced End-Of-Course (EOC) tests for Algebra 1 and 2, Geometry, Biology, Chemistry, Physics, Physical Science, US History, Social Science and English 1 between 1988 and 1991. Except for a four year interlude in which some tests were made a local option, all students taking these courses were required to take the state tests. Easier versions of these courses not assessed by a state test do not exist, so virtually all North Carolina high school students take at least six of these exams. Test scores are reported separately on the student's transcript. Most teachers have been incorporating EOC exam scores into their course grades and

a state law now mandates that, starting in the year 2000, the EOCE test scores must have at least a 25% weight in the final course grade.

How are North Carolina and New York doing? Did student test scores go up in North Carolina after they implemented their end-of course exams. Yes they did. In fact according to Grissmer, Flanagan, Kawata and Williamson (2000), 4th and 8th grade test scores rose more rapidly from 1990 to 1996 in North Carolina than in any other state.[28] While suggestive, such a finding is not conclusive. North Carolina was introducing other accountability policies--rewards for school improvement and sanctions for poor performance--at the same time, so the increase in 8th grade test scores could be due to these efforts not the CBEEES.

New York has had the Regents exams for more than one hundred years, so there is no reason to expect particularly rapid test score gains. The effects of the Regents exam system can be studied by examining cross section data as was done in the international and Canadian studies described above.

Effects on Peer Culture: The Educational Excellence Alliance survey on the attitudes and behavior of 35,000 students in 135 schools in New York, Connecticut, Massachusetts and New Jersey provides a good data set for testing whether CBEEES tend to generate a student peer culture that is more supportive of learning and has a more favorable perception of teachers.[29] About 60 percent of the schools surveyed were not in New York State.[30] For each of the 135 schools surveyed we calculated the proportion of students who answered each question affirmatively by gender and by grade. Multiple regression models were estimated predicting these proportions as a function of gender, grade, parental education of the students in the school, proportion of students living in single parent homes, proportion of students Hispanic, proportion of students Asian and proportion of students African-American, a dummy variable for non-public school and dummy variables for state.

The findings are quite interesting. Attitudes toward teachers were more positive in New York. When students were asked what motivated them to study hard, New Yorkers were 30 percent more likely to respond "to please or impress my teacher," 17 percent more likely to say 'my teachers encourage me to work hard.' and 14 percent more likely to say "the teacher demands it." New York students were also significantly more likely to say "my teachers grade

me fairly", "my teachers maintain good discipline in the classroom" and that classes are "interesting."

The peer culture was also better. New York students were 10 percent more likely to say "My friends think it is important for me to do well in [science, math, English] at school." They were nearly 25 percent more likely to be annoyed when "other students talk or joke around in class" or "try to get the teacher off tract." In addition New York students were significantly more likely to say they were motivated by a desire to learn the material and more likely to report they were interested in what they were studying and more likely to talk with their friends outside of class about what they were studying.

The better attitudes translated into better behavior. New York students spent significantly more time studying for history exams, more time doing homework and did a larger share of the homework that was assigned. They also paid closer attention in class and contributed to class discussion more frequently.

Impacts on Learning: New York's students are more disadvantaged, more heavily minority and more likely to be foreign born than students in most other states. Consequently, when one compares student achievement levels, family background must be taken into account. Considering the high incidence of at-risk children, New York students do remarkably well. Table 3 presents the results of a linear regressions predicting 1992 NAEP math scores and 1991 mean SAT-M + SAT-V test scores for all states for which data are available. With the exception of the dummy variable for New York State, all right hand side variables are proportions--generally the share of the test taking population with the characteristic described. In the analysis of 8th grade math scores the controls for student background were: the proportion of people under age 18 who live in poverty, a schooling index for the adult population, percent foreign born, percent public school students who are black and percent public school students who are Hispanic, parent's education, the poverty rate, percent black and percent foreign born all had significant effects on math achievement in the expected direction. New York State's mean NAEP math score was a statistically significant 9.6 points (or about one grade level equivalent) above the level predicted by the regression model.

[Table 3 about here]

In the analysis of SAT test score means, the control variables were a parents' education index, percent black, percent in private schools, percent in large schools, percent who had taken 3 or more courses in math and English and the percent of high school graduates who take the SAT. New

Yorkers did significantly better (46 points better) on the SAT than students of the same race and social background living in other states. For individuals the summed SAT-V + SAT-M has a standard deviation of approximately 200 points. Consequently, the differential between New York State's SAT mean and the prediction for New York based on outcomes in the other 36 states is about 20 percent of a standard deviation or about three-quarters of a grade level equivalent (Bishop, Mane and Moriarty 2000).

Further evidence on the effect of New York's CBEEES system comes from an analysis of test score gains between 8th and 12th grade that appeared recently in the Brookings Papers on Education Policy. The results of our analysis of NELS:88 data are presented in Figure 2. We found significantly larger test score gains (about 40 % of a grade level equivalent) by students in New York (Bishop, Mane, Bishop and Moriarty 2001). Increases in the number of courses required to graduate and minimum competency exams did not have significant effects on test score gains.

This paper also analyzed 1996 and 1998 state cross section data on 8th grade NAEP reading, mathematics and science test scores. Our models included controls for the following demographic characteristics of the students attending school in the state: the share of children living in poverty, parental education, the share of public school students who are African-American, the share who are Hispanic and the share who are Asian-American. States that have moderate or high stakes tests for students tend to have also adopted school accountability systems that reward high achieving schools or sanction failing schools that do not improve during the early 1990s. This means that unbiased estimates of the effect of minimum competency exams and CBEEES are possible only when the presence or absence of other standards-based reform initiatives is taken into account. We, therefore, studied the impact of four different policies:

1. Rewards for schools that improve on statewide tests or exceed targets set for them
2. Sanctions for failing schools—closure, reconstitution, loss of accreditation etc.
3. Minimum competency exams
4. Moderate Stakes Curriculum-based External Exit Exam System—i.e. the New York/North Carolina stakes for students policy mix during the 1990s[31]

Results of the analysis are presented in Figure 3. The policy that clearly had the biggest effects on test scores is the moderate-stakes curriculum-based external exit exam system. In science and mathematics 8th graders in New York and North Carolina were one-half of a grade

level equivalent (GLE) ahead of comparable students in states without minimum competency exams or CBEEES. They were also a 63 percent of a GLE ahead in reading.

Stakes for teachers and schools also had significant effects on all three measures of 8th grade achievement. Students living in states that in 1996/7 were both rewarding successful schools and threatening to sanction failing schools were about 28 percent of a GLE ahead in all three subjects of students in states that did neither. Public reporting is necessary for the implementation of these other policies but on its own it had no discernable effect on student achievement. Point estimates for the impact of minimum competency exams were positive but small and only one of the three coefficients was significant at the 10 percent level on a one tail test.

Effects of Requiring Students to Pass Regents Exams to Graduate from high school:

Until quite recently Regents courses and Regents exams were voluntary. Forty to fifty percent of students avoided them by electing to take watered down 'local' classes instead either to reduce their work load or to boost their GPA. Concern grew that many of these local courses were a waste of time. The Board of Regents decided to require higher standards by introducing more demanding 4th and 8th grade assessments and by eliminating the local course option in five core academic subjects. New Regents exams were developed and beginning with students entering 9th grade in 1996 the diploma was awarded only to students who were able to get a 55 or better on the new six-hour English exam. The requirement to take and pass exams in five subjects applies to those entering 9th grade in 1999 or later.

How has student achievement fared during the five-year period since the Regents announced that graduating from high school would be made conditional on passing five Regents exams. AP course taking, SAT scores and college attendance rates are all up. The proportion of high school graduates meeting the requirements for a 'Regents diploma' (a 65 or better on eight Regents exams) rose from 40 percent in 1995 to 49 percent in 2000. The data on the proportion of New York students that passed with 65 percent or better is presented in Figure 4. The number of students getting a 65 or better on Regents exams increased by 52 percent in English, 43 percent in mathematics and 46.5 percent in global history. During that same period the number of high school diplomas awarded in New York rose 7 percent. The high school completion rate--the ratio of high school diplomas awarded to fall enrollment in 8th grade five years earlier--fell only slightly from .742 to .728. The success of New York students in meeting the new higher

standards is a good sign that Michigan's effort to raise standards may also succeed.

The drive for higher standards in New York high schools has apparently achieved some initial success. Much greater challenges lie ahead, however. The graduating class of 2000 was the first cohort required to pass a Regents English exam. The Class of 2001 must pass both the English and the mathematics exam. Failure rates on the math exam are much higher, particularly in the state's urban school systems. Students graduating after 2003 will face the even tougher challenge of passing five exams and graduation rates may decline even more. New York's high non-completion rate shouldn't become a problem in Michigan because passing the MEAP HST is not a graduation requirement.

Michigan--Effects of the Merit Award program during the first two years:[32] What does the behavior of the first two cohorts of students eligible for Michigan Merit Awards, the class of 2000 and the class of 2001, tell us about the effects of the program. Table 4 presents data collected from the Michigan Department of Education, the State Budget Office and the Merit Award program tracking the number of people taking and passing MEAP HST exams. The first two columns of the table report the number of students who took and passed the MEAP HST during the spring of 1998 and 1999. Governor Engler proposed the Merit Award program in January 1999, four months before the high school juniors were supposed to take the HST test in May. While the authorizing legislation didn't pass until a couple of months later, passage was expected throughout the spring and most students were aware that taking and passing all the HST tests would probably result in their getting a $2500 scholarship. Consequently, the very large (11,316) increase between 1998 and 1999 in the numbers taking the reading HST was at least in part due to the announcement of the program. By my calculation the proportion of 12th graders taking the test during the spring of their junior year rose from 75 percent to 85 percent.

The third column of the table gives counts for the graduating class of 2000. Most of these students first took the HST in spring 1999. The 10,473 increase in numbers taking the reading HST are seniors who took the HST in either fall 1999 or Spring 2000 after not taking the test in spring 1999. The proportion of 11th graders in the class of 2000 cohort taking the HST reached 95.7 percent.[33] This number would almost certainly have been a lot smaller in the absence of the Merit Award program. The fourth column of the table gives counts for the graduating class of 2001. The next year, however, number of students from the class of 2001

taking the MEAP HST reading test dropped by 3,691. The ratio of test takers to 11th graders in the Class of 2001 was slightly lower, 93.3 percent.

More important than the increase in the number of students taking the HST exams, has been the big increase in the number of students demonstrating that they met or exceeded Michigan's education goals. In reading, for example, the number of students meeting standards increased by 13,733 between spring 1998 and spring 1999 and then by another 8032 by the time the class of 2000 had completed senior year. Despite a small decline in the size of the senior class to 2001, the number of graduates meeting Michigan standard increased by 1690 in reading, 1443 in mathematics, 2560 in science and 6714 in writing.

Even though the share of the high school class taking the HST test was increasing substantially, the proportion of test takers meeting the goals went up significantly. For math the proportion meeting the standard rose from 60.5 percent in 1998 to 68.4 percent for the class of 2001. For reading the proportion rose from 58.9 percent in 1998 to 74.2 percent in 2001. The proportion passing in science rose from 51.7 percent to 60.0 percent. The proportion meeting the writing standard rose form 56.6 percent to 68.5 percent.

Effects on College Attendance Rates: The Hope Scholarship program in the State of Georgia significantly increased college attendance rates particularly at colleges in the state.[34] The Hope scholarship, however, is considerably more generous than the MEAP Merit Award, so the impact of the Merit award is likely to be a lot smaller. The Hope Scholarship lasts four years (if you keep your average above B in college), pays full tuition at Georgia public universities and community colleges and a similar amount at private colleges in Georgia and pays nothing for students attending out of state. The Merit Award is for one year only, includes students attending college out-of-state and the $2500 award covers only about half of Michigan State's resident tuition of $4879. and about four-fifths of tuition at schools like Eastern Michigan and Western Michigan University. Consequently one would expect a much smaller response to the Merit Award than to the Hope scholarship and a response that is not focused on inducing students to attend college instate.

Did college attendance go up? Data is not available on the college attendance rate of students in the Class of 2000 or 2001. There is data, however, on trends in the number of students in each graduating class who took the ACT test and how they did on the test. Since the ACT test is the college admissions test used by almost all Michigan colleges and universities, the

count of students taking this test is a good indicator of achievement at the end of high school and of trends in the number of students expecting to go to college. These data are presented in Table 5 and table 6. ACT's estimate of the share of graduating classes taking their tests is presented in column 5 of Table 5. These rates were stable from 1997 to 1999. Classes graduating in 1999 and earlier years were not eligible for Merit Awards. The Merit Award kicked in with the class of 2000 and the ACT test-taking rate increased 2 percentage points to 71 percent for that class and then fell back to 69 percent in 2001. Furthermore, the share of ACT test takers from minority groups also rose in 2000. This suggests that the Merit Award may have stimulated a larger proportionate increase in college going among minority groups than among whites. Mean ACT test scores were stable. This suggests that the increase in the proportion of seniors taking the ACT since 1994 did not lower the average test scores of ACT test takers.

The ACT data on the proportion of the graduating class taking the ACT has problems, however. The denominator of the ratio, the number of public plus private high school graduates, is a projection made in 1998 (using 1996 as a baseline) by the Western Interstate Commission on Higher Education. Actual data on the number of graduates or seniors at public and private high schools would be preferable, but annual data is not available for private high schools. Consequently, we present two other indicators of college going plans--the ratio of unduplicated ACT test takers to fiscal-year equivalent (FYE) public regular high school and charter school seniors in column 6 of Table 5 and the ratio of ACT test takers to FYE 11th graders lagged one year in row 3 of table 6. Both of these indicators suggest that ACT test taking rates rose in 2000 and stayed up in 2001. There was about a 2 percent increase in ACT test taking rates in 2000, the year the Merit award began. This is consistent with the Merit award boosting college attendance but far from conclusive because something else might have caused the increase and college enrollment trends might be different from ACT test taking trends.

The fifth row of Table 6 presents data on undergraduate and graduate enrollment in Michigan's four-year public colleges and universities for fall 1998 through Fall 2001 obtained from the State Budget Office.[35] There seems to have been an increase in the growth of four-year public college enrollment in the year the MEAP Award starts (see row 6). However, this does not take account of the growing size of the pool of recent high school completers. How does enrollment compare to the stock of high school seniors who have completed high school in the last six years? That ratio is presented in row 7 and rates of growth for the enrollment ratio are in

row 8 of Table 6. Here again there appears to be a one-year acceleration in the growth of public four-year college enrollment in the year 2000. The effect is small, however, and might be a chance event or a consequence of Michigan students choosing to attend college instate. Conclusive evidence on the impact of the merit award program on college attendance rates in Michigan must wait until more data becomes available. Data is needed on first-time enrollment rates and total undergraduate enrollment in Michigan and neighboring states for a number of years both before and after 2000. This means waiting a couple of years until IPEDS data becomes available.

V. Conclusion

I conclude that the case for the Michigan Merit Award program is very strong. Many, probably most, of the Michigan's secondary school students have not been devoting as much time and energy to learning as their parents and the public would like. Students who blow off high school pay a very high price; a much larger price than they imagine when they are in school. They imagine they will be able to go to college regardless of low grades, regardless of low achievement. But, in fact, their chances of completing a degree program are almost zero. They are also unaware that applying themselves in high school helps them get jobs that offer training and promotion opportunities and eventually higher wage rates. Consequently, it is sensible for state government to purposely try to strengthen incentives to study and to make them absolutely transparent to students and parents. That is exactly what the Merit Award program accomplishes.

The Merit Award's use of the MEAP HST as the primary method for selecting scholarship winners creates a moderate-stakes curriculum-based external exit exam system in state of Michigan. Experience with similar examination systems in Canada, New York and North Carolina is very positive. Michigan can reasonably anticipate that the Merit Award program will increase student effort and learning, make parents stronger advocates of higher standards, increase college attendance and reduce college drop out rates.

Table 1: Time Use By Students

	Hours Watching T.V per Week		Reading Time per Week
	Students	Adults	Students
U.S.	19.6	15.9	1.4
Austria	6.3	10.6	4.9
Canada	10.9	13.3	1.5
Finland	9.0	9.0	6.0
Netherlands	10.6	13.4	4.3
Norway	5.9	7.2	4.3
Switzerland	7.7	9.0	4.8

Source: Hours spent per week on each activity derived from time diary studies. Organization of Economic Cooperation and Development, Living Conditions in OECD Countries, 1986, Tables 18.1 & 18.3.

Table 2: Impact of Literacy and Schooling on the Earnings and Unemployment of Males

Prose Literacy	Earnings	Unemployment Rate--1992	**Schooling**	Earnings	Unemployment Rate--1992
Level 1	$48,965	2.3 %	BA or more	$38,115	4.8 %
Level 2	$39,941	4.1 %	Assoc. Degree	$31,855	5.5 %
Level 3	$29,610	6.4 %	13-15 Yrs	$27,279	7.4 %
Level 4	$22,046	11.5 %	12 Yrs	$22,494	8.2 %
Level 5	$15,755	14.9 %	9-11 Yrs	$16,194	12.4 %

Source: National Adult Literacy Survey of 1992, National Center for Education Statistics, Literacy in the Labor Force,

Table 3: New York State Student Achievement compared to other States in the early 1990s

	NYS	Partici-pation Rate	Parents Educ. Index	Prop. Black	Prop. Hispanic	Prop. Foreign Born	Prop. In Poverty	Prop. Private School	Prop. Large School	Prop. 3+ Math Courses	Prop. 3+ Eng. Courses	R Sq/ RMSE	Mean/ Std Dev
1992 NAEP Math 8th Grade	9.6 (2.1)**		68** (2.7)	-32*** (6.1)	-1 (.1)	-66*** (3.2)	-.52** (2.5)					.831 4.23	
Total SAT	46** (2.7)	-.68** (2.6)	370*** (6.4)	-.135 (3.2)				60 (1.6)	-44* (1.8)	85 (1.3)	-36 (.3)	.926 14.8	925 55
SAT Independent Variables--Mean	.027	.414	.581	.078				.207	.120	.617	.797		
Std. Deviation	.164	.240	.097	.064				.082	.113	.067	.038		

Source: Analysis of 1991 state average scores on the NAEP mathematics assessment and the summed math and verbal SAT-I tests. The mean and standard deviation for the variables used in the SAT analysis are in rows 5 and 6 of the table. NAEP test results are reported on a scale where a grade level equivalent is about 10 points.

*** $p < .01$ on a two tail test
** $p < .05$ on a two tail test
* $p < .10$ on a two tail test

Table 4: Michigan Public School Students Who Met or Exceeded State Standards on the MEAP High School Tests

		Spring 1998	Spring 1999	Class of 2000	Class of 2001	Class of 2002
Math	Endorsed (% of Tested)	**60.5 %**	**63.6 %**	**64.8 %**	**68.4 %**	
	Number Endorsed	43,122	53,632	59,592	61,035	
	Increase from Prev. Yr		10,510	**5,960**	1,443	
Reading	Endorsed (% of Tested)	**58.9 %**	**67.3 %**	**69.4 %**	**74.2 %**	
	Number Endorsed	42,216	55,949	63,981	65,671	
	Increase from Prev. Yr.		13,733	**8,032**	1,690	
Science	Endorsed (% of Tested)	**51.7 %**	**51.0 %**	**55.6 %**	**60.0 %**	
	Number Endorsed	36,559	41,911	50,723	53,283	
	Increase from Prev. Yr.		5,352	**8,812**	2,560	
Writing	Endorsed (% of Tested)	**56.6 %**	**52.5 %**	**58.4 %**	**68.5 %**	
	Number Endorsed	39,104	41,868	51,608	58,322	
	Increase from Prev. Yr.		2,764	**9,740**	6,714	
# Public School Students who took the High School Test in Reading		70,401	81,717	92,190	88,499	
Public High School Pub. FYE 12th Graders for this MEAP cohort		94,342	96,293	96,293	94,837	
Ratio MEAP HST takers to FYE Public High School 12th Graders		74.6%	84.9%	95.7%	93.3%	
# Students Eligible for Merit Award		---	---	43,068	48,760	

Sources: Data for 1998 and 1999 are for first-time test takers in the Spring administration of the MEAP HST. "MEAP Scores Reflect At Least 20,000 Students Eligible for Merit Scholarships" at www.meritaward.state.mi.us/whatsnew/newsrel/1999/092899_2.htm. The data for the Class of 2000 is for the graduating class of 2000 and represents the highest test score for students who had multiple opportunities to take the test before graduating in 2000. It is from www.meritaward.state.mi.us/merit/meap/results/data/2000summary.htm. Data on the class of 2001 is from http://www.meritaward.state.mi.us/mma/results/class01/summary.htm. Audited data on fiscal year equated students enrolled in 12th grade in Michigan regular public and charter high schools was provided by the Michigan State Budget Office.

BEST COPY AVAILABLE

Table 5: Trends in ACT Test Taking and Scores for Michigan

Grad. Class	Public H.S. Grads	FYE Public H.S. Seniors Prev Fall	All Grad. Class taking ACT	ACT/ Grads	ACT/ Public H.S. Seniors	ACT All Stud-ents	ACT White Students	ACT Black Students	ACT Hispanic Students	ACT Asian Student	Minority Students Taking ACT
1991	88,234	91,769					21.2	16.9	19.3	22.4	
1992	87,756	90,332	62,206		**.689**		21.2	16.9	19.0	22.1	
1993	85,302	89,285	62,035		**.695**		21.4	17.0	19.2	22.4	
1994	83,385	87,498	60,401	63 %	**.690**	21.0	21.6	17.2	19.2	22.5	
1995	84,628	88,143	62,416	64 %	**.708**	21.1	21.6	17.2	19.3	22.7	
1996	85,530	87,826	62,952	64 %	**.717**	21.1	21.7	17.2	19.5	22.4	7601
1997	89,695	90,348	66,628	68 %	**.738**	21.3	21.8	17.2	20.0	22.6	7898
1998	92,732	92,690	68,769	68 %	**.742**	21.3	21.9	17.1	19.3	22.6	8510
1999	94,125	94,342	70,669	69 %	**.749**	21,3	21.9	17.1	19.6	22.3	8743
2000	*94,710*	96,293	73,918	71 %	**.768**	21.3	21.9	17.0	19.8	22.6	9192
2001	*93,260*	94,837	72,450	69 %	**.764**	21.3					
2002		96,612									

Source: Data in column 2 on the number of public high school grads is from Table 24 of Projections of Educational Statistics to 2011, NCES 2001-083. The numbers in parenthesis for 2000 and 2001 are projections made by NCES. Data on ACT test taking is from "ACT High School Profile Report--H.S. Graduating Class 2000: Michigan" and "1996 Progress Towards the National Education Goals" at www.mde.state.mi.us/reports/neg/neg1996/goal3.shtml. ACT mean and % of HSG taking is from www.act.org/news/data.html. Column 5, the % of graduates taking ACT, is calculated by dividing the number of ACT takers graduating from Michigan high schools that year by a projection of the number of high school graduates (both Public and private) made by the Western Interstate Commission on Higher Education, Knocking at the College Door,1998. The projection was made in 1998. Annual data on seniors or graduates at private high schools was not available and data on the actual number of graduates of Michigan public high school available from NCES for 2000 and 2001 were clearly unreliable. Consequently, the denominator of column 6 is the actual number of fiscal year equated students in the 12th grade of regular public high schools and charter schools kindly supplied by Mechelle Marcum of the State Budget Office.

Table 6: Trends in ACT Test Taking and College Attendance for Michigan

Year	1998	1999	2000	2001	2002
Public High School head count of 11th Graders for graduating class	**102,613**	**102,991**	**105,706**	**104,630**	
Students taking the ACT in Graduating Class	68,769	70,669	73,918	72,450	
Ratio All ACT takers/Public FYE 11th graders	67.0%	68.6%	69.9%	69.2%	
Mean ACT Score	21.3	21.3	21.3	21.3	
Number of Students in Public 4 year Colleges—Fall Head Count	**264,527**	**268,441**	**275,651**	**279,969**	
% growth of enrollment from previous year		1.5%	2.7%	1.6%	
Ratio-Public 4 year College Enrollment to Sum of FYE 12th graders in past 6 years	49.4%	49.6%	50.2%	50.3%	
% growth of enrollment ratio from t-1		0.5%	1.0%	0.3%	

Sources: ACT data is from "ACT High School Profile Report--H.S. Graduating Class 2000: Michigan" and "1996 Progress Towards the National Education Goals" at www.mde.state.mi.us/reports/neg/neg1996/goal3.shtml and by personal communication from ACT Research office. Data on FYE students enrolled in 11th and 12th grade in Michigan public high schools was provided by the State Budget Office. Data on the number of students enrolled in Michigan's public 4-year colleges was kindly provided by Glen Preston of the State Budget Office. We use data on undergraduate enrollment plus graduate enrollment because separate data on undergraduates or freshman are not available for 2000 and 2001. Data on total enrollment including non-resident enrollment because some universities changed their definition of 'resident' student during this period.

References

Bishop, John. "Is the Test Score Decline Responsible for the Productivity Growth Decline." American Economic Review, Vol. 79, no. 1, March 1989a.

Bishop, John "The Productivity Consequences of What is Learned in High School." Journal of Curriculum Studies, 1989b.

Bishop, John and Ferran Mane. (2001) "Incentive Effects of New York State's Reform Strategy" to appear in Parental Choice vs. Best Systems: What Improves Education eds. Margaret Wong and Herbert Walberg, Lawrence Earlbaum Associates, 1-35 .

Bradley, Ann. "Not Making the Grade: Teacher firing spurs debate over standards and expectations for students." Education Week, Sept. 13, 1993, p. 1, 19-21.

Cooper, Harris M. Homework. White Plains, New York: Longman, 1989.

Fleming, M. and Chambers, B. (1983) Teacher-made Tests: Windows on the Classroom. San Francisco: Jossey Bass.

Frederick, W. C. "The Use of Classroom Time in High Schools Above or Below the Median Reading Score." Urban Education 11, no. 4 (January 1977): 459-464.

Frederick, W.; Walberg, H.; and Rasher, S. "Time, Teacher Comments, and Achievement in Urban High Schools." Journal of Educational Research 73, no. 2 (November-December 1979): 63-65.

Goodlad, J. A Place Called School. New York: McGraw-Hill, 1983.

Grissmer, David and Ann Flanagan, Jennifer Kawata and Stephanie Williamson. Improving Student Achievement: What NAEP test scores tell us. Rand Corporation, 2000, 1-271.

Hart (Peter D.) Research Associates, (1995) "Valuable Views: A public opinion research report on the views of AFT teachers on professional issues." Washington D.C.: American Federation of Teachers, 1995, 1-24.

High School and Beyond. Data File Users Manual. National Opinion Research Corporation, Chicago, Illinois, 1982.

Hollenbeck, K., and Smith B. The Influence of Applicants' Education and Skills on Employability Assessments by Employers. Columbus: The National Center for Research in Vocational Education, The Ohio State University, 1984.

International Assessment of Educational Progress, Learning Science. Princeton, New Jersey: Educational Testing Service 1992a.

International Assessment of Educational Progress, Learning Mathematics. Princeton, New Jersey: Educational Testing Service 1992b.

Jencks, Christopher and Crouse, James. "Aptitude vs. Achievement: Should We Replace the SAT?" The Public Interest, 1982.

Juster, Thomas and Stafford, Frank. "The Allocation of Time: Empirical Findings, Behavioral Models and Problems of Measurement." Ann Arbor, Mich.: Survey Research Center, Institute for Social Research,

Longitudinal Survey of American Youth. "Data File User's Manual" Dekalb, Ill: Public Opinion Laboratory, 1988.

National Center for Education Statistics. The Digest of Education Statistics:. [various years] Wash. D.C.: US Department of Education..

Organization of Economic Co-operation and Development. Living Conditions in OECD Countries: A Compendium of Social Indicators. Social Policy Studies No. 3. Paris, France: Organization for Economic Co-operation and Development, 1986.

Organization of Economic Co-operation and Development. Education at a Glance. Paris, France: Organization for Economic Co-operation and Development, 1992.

Powell, Arthur; Farrar, Eleanor and Cohen, David. The Shopping Mall High School. New York, New York: Houghton Mifflin, 1985.

Rohwer, W.D. and Thomas, J.W. (1987) "Domain specific knowledge, cognitive strategies, and impediments to educational reform." in M. Pressley (Ed.), Cognitive Strategy Research. New York: Springer-Verlag.

Rubenstein, Ross and Gary Henry, "Paying for Grades: Impact of Merit-Bas3ed Financial Aid on Educational Quality" Georgia State University, Oct. 2000, 1-30.

Sizer, Theodore R. Horace's Compromise: The Dilemma of the American High School. Boston: Houghton Mifflin, 1984.

Steinberg, Laurence. Beyond the Classroom. New York: Simon and Schuster, 1996, 1-223.

Thomas, John W. "Expectations and Effort: Course Demands, Students Study Practices and Academic Achievement." paper presented at a 1991 Conference on Student Motivation sponsored by the Office of Educational Research and Improvement.

Tim, Paper for ILR 360, October 1986, 1-6.

Wexler, Philip. Becoming Somebody: Toward a Social Psychology of School. Washington, D.C.: Falmer Press, 1992, 1-163.

Wiley, David E. "Another Hour, Another Day: Quantity of Schooling, a Potent Path for Policy." In Schooling Achievement in American Society, edited by William H. Sewell, Robert M. Hauser, and David L. Featherman. New York: Academic Press, 1976.

Zoch, Paul. "A Teacher's View of Accountability." Selected Readings on School Reform. Winter 1999, Vol. 3, No. 1, Washington, DC: Thomas B. Fordham Foundation. 70-71

Appendix A: Description of the Michigan Merit Award Program
[http://www.meritaward.state.mi.us/mma/faq.htm]

GENERAL INFORMATION

33. What is the Michigan Merit Award Program?

Public Act 94 of 1999 (the *Michigan Merit Award Scholarship Act*) provides for a merit-based program for high school seniors to reward student achievement and to make postsecondary education more affordable. Beginning with the Class of 2000, students who meet certain criteria will be eligible for a Michigan Merit Award of $2,500 to be used at any approved postsecondary educational institution. Beginning with the Class of 2005, there will be an additional potential award of up to $500.

34. Is the Michigan Merit Award available to all Michigan students?

Yes. The Michigan Merit Award is available to all Michigan students (including public school, public school academy, nonpublic school, and home schooled students) who meet all eligibility requirements. This also includes Michigan residents attending a high school out of state or out of the country..

35. When did the Michigan Merit Award begin?

The Michigan Merit Award was first available to students of the high school graduating Class of 2000 who met all the eligibility requirements.

36. Who administers the Michigan Merit Award Program?

The Michigan Merit Award Board is established within the Michigan Department of Treasury to administer the program. The Michigan Merit Award Board consists of the State Treasurer, the Superintendent of Public Instruction, the Director of the Department of Career Development, and four public members appointed by the Governor. The Michigan Merit Award Board is responsible for developing the rules and processes by which the program will be implemented and administered.

37. What are some of the primary eligibility requirements for receiving the Michigan Merit Award?

The Michigan Merit Award Act sets forth general eligibility requirements. The Michigan Merit Award Board is responsible for establishing specific eligibility requirements and determining whether they have been met.

To be eligible, a student must take the MEAP High School Tests in mathematics, reading, science, and writing. Students who score at Level 1 (Exceeded Michigan Standards) or Level 2 (Met Michigan Standards) on these four tests and meet all other eligibility requirements will qualify to receive a $2,500 Michigan Merit Award.

For a student who takes all four of the above-specified subject tests, meets or exceeds state standards on at least two, and meets all other eligibility requirements, there are two alternate ways to qualify:

- *Alternate A*: The student also scores in the 75th percentile or above on the ACT or SAT.
- *Alternate B*: The student also achieves qualifying scores on the ACT WorkKeys job skills assessment tests as determined by the Michigan Merit Award Board.

Please note that under both *Alternates*, the student must take all four of the above-specified MEAP subject tests and achieve Level 1 or Level 2 on at least two.

38. What are some of the other eligibility requirements for receiving the Michigan Merit Award?

The student must have graduated from high school or passed the General Educational Development (G.E.D.) test. The student must be enrolled in an approved postsecondary educational institution. There is also a requirement that the student must not have been convicted of a felony involving an assault, physical injury, or death.

39. Are there other situations not covered by this Q&A document?

BEST COPY AVAILABLE

Yes. Space does not permit us to discuss in detail every conceivable circumstance. This Q&A document reflects existing law (Public Act 94 of 1999) as of January 2002. If you have questions about some aspect or circumstance not specifically covered by this document, please do not hesitate to call us toll-free (888)95-MERIT.

7th & 8th GRADERS

40. How does the Michigan Merit Award program apply to 7th and 8th graders?

Starting with the high school graduating Class of 2005, students have an opportunity to earn up to $500 in addition to the $2,500 they can earn from the Michigan Merit Award, bringing the total possible Michigan Merit Award to $3,000. This award program is based on a student taking all four MEAP tests (Math, Science, Reading and Writing) offered in 7th and 8th grade, and meeting or exceeding state standards on at least two of the four tests.

Meet/Exceed State Standards on Earn an extra

2 of the 4 tests $250

3 of the 4 tests $375

All 4 tests $500

41. Can a student who did not take the 7th grade MEAP tests (or 8th grade MEAP tests) qualify for the additional monies?

No. A student must take the 7th grade MEAP tests during their 7th grade year, and the 8th grade MEAP tests during their 8th grade year. A student is not eligible if they only take the MEAP tests offered in 7th grade, or only the MEAP tests offered in 8th grade. They must take all four tests in the appropriate years offered.

42. Are retakes offered?

No. Retakes for the 7th and 8th grade MEAP tests are not offered.

43. How will 8th graders and parents be notified?

We strongly encourage parents to stay in touch with the Guidance Office at their child's middle school which maintains student test scores. The Michigan Merit Award office is currently modifying its database system which will allow us to notify 8th grade students of their eligibility for these additional funds. However, this notification process is still under development.

44. When is the award available?

The additional award will be made available to students upon graduation from high school (Class of 2005 and after). The student must first qualify at the high school level for the Michigan Merit Award ($2500) in order to receive any additional monies ($250, $375 or $500). Funds will be paid in accordance with Michigan Merit Award procedures.

45. My child does not attend a public school. What are our options?

Students who do not attend a public school may test at their local public school, at their own nonpublic school, or at a MEAP Test Center.

ELIGIBLE COSTS

26. May the Michigan Merit Award be used for any expense associated with my attendance at an approved institution?

The Michigan Merit Award may be used for "eligible costs" as determined by the Michigan Merit Award Board.

27. What are "eligible costs?"

The following costs will be considered eligible costs for purposes of the Michigan Merit Award:

- Tuition and fees, including cost of rental or purchase of equipment or materials required of all students in the same course of study

- Books, supplies, miscellaneous educational expenses, transportation, rental or purchase of a personal computer
- Reasonable room and board
- Dependent care during periods that include, but are not limited to, class time, study time, field work, internships and commuting time
- Disability-related expenses, including special services, personal assistance, materials, equipment and supplies
- Educational costs associated with a cooperative education program

The procedure to use Michigan Merit Award funds is simple: we pay the institution on your behalf. In all likelihood, your institution will apply your Michigan Merit Award funds toward primary costs, enabling you to use your own money (which otherwise would go to pay for primary costs) for other expenses such as a new computer, etc. Talk with your institution's Student Financial Aid Office to determine how your award funds may be accessed. Each institution is different and has its own policies. We do not reimburse students for purchases.

28. What if I already have a scholarship?

Talk with your institution's Student Financial Aid Office if you anticipate receiving another scholarship. There may be preferred ways to accept the Michigan Merit Award and avoid unnecessary financial entanglements. For example, in some circumstances, student athletes could jeopardize their NCAA status if they accept a Michigan Merit Award. They may wish to consider waiting to use their Michigan Merit Award.

The Michigan Merit Award may be used for eligible costs. Often "full-ride" scholarships include direct education-related costs such as tuition, fees, room, board, and books, but not indirect education-related costs. In such cases, a student may choose to use the Michigan Merit Award to pay for indirect education-related costs such as materials, supplies, day care, miscellaneous educational expenses, or a computer. Talk with your institution's Student Financial Aid Office.

Because students have up to seven years to utilize the award, some students may decide to apply their Michigan Merit Award toward graduate studies.

OTHER ISSUES

29. May I use the Michigan Merit Award for summer school?

Yes, but not for the summer immediately after high school graduation. For example, if you graduated from high school in June 2002, we will not pay for summer 2002 classes. Your first payment cannot be made any sooner than for classes taken in fall 2002.

30. May I use the Michigan Merit Award for graduate school?

Yes. You may apply your Michigan Merit Award toward a graduate program at an approved institution, provided that you start the graduate program no later than seven years after the date of your graduation from high school.

31. Is the Michigan Merit Award affected by my high school grade point average?

No. The Michigan Merit Award is based solely upon your scores on the MEAP (and, if necessary, ACT, SAT and/or WorkKeys tests).

32. Is the amount of my Michigan Merit Award based on my family's income?

No. The Michigan Merit Award is based solely upon your scores on the MEAP (and, if necessary, ACT, SAT and/or WorkKeys tests).

34. Is the Michigan Merit Award available to all Michigan students?

Yes. The Michigan Merit Award is available to all Michigan students (including public school, public school academy, nonpublic school, and home schooled students) who meet all eligibility requirements. This also includes Michigan residents attending a high school out of state or out of the country.

Appendix B: Design and Validity of the MEAP Test

[excerpt from MEAP web site]

An Overview

The Michigan Revised School Code (1977) and the State School Aid Act (1979) require the establishment of educational standards and the assessment of students' academic achievement. Accordingly, the State Board of Education, with the input of educators throughout Michigan, approved a system of academic standards and a framework within which local school districts could develop and implement curricula as they see fit.

The Michigan Educational Assessment Program (MEAP) tests were developed for the purpose of determining what students know and what students are able to do, as compared to these standards, at key checkpoints during their academic career. Hundreds of educators from throughout Michigan continue to be involved in the development and ongoing improvement of these tests. No other tests measure what is expected of Michigan students, nor measure the performance of Michigan students against established academic standards.

The MEAP tests have been recognized nationally as sound, reliable and valid measurements of academic achievement. Students who score high on these tests have demonstrated significant achievement in valued knowledge and skill. Further, the tests provide a common denominator to measure how well students are doing, and to assure that all Michigan students are measured on the same skills and knowledge, in the same way, at the same time.

Properly used, the MEAP tests can:

- Measure academic achievement as compared to expectations, and whether it is improving over time;
- Determine whether improvement programs and policies are having the desired effect;
- Target academic help where it's needed.

Admittedly, there is some pressure associated with taking the MEAP tests, but it is a positive pressure. Competitive scholastic experience provides Michigan students with excellent preparation for the real world which awaits them after high school graduation, and helps assure that they possess the knowledge and skill necessary for a successful future.

The Michigan Educational Assessment Program is all about effort, improvement and academic excellence. Michigan students are expected to learn and grow, and the MEAP continues to make a valuable contribution in providing them the opportunity to measure their academic progress. The MEAP tests and administration of the tests are far from perfect, but our collective effort should be student focused with a clear bias toward accurate analysis, constructive criticism and continual improvement.

Purpose of the MEAP

The MEAP tests were developed to measure what Michigan educators believe all students should know and be able to achieve in five content areas: mathematics, reading, science, social studies, and writing. The test results paint a picture of how well Michigan students and Michigan schools are doing when compared standards established by the State Board of education. The MEAP test is the only common measure given statewide to all students. It serves as a measure of accountability for Michigan schools.

Results of MEAP tests can be used by schools for school improvement purposes. The results indicate overall strengths and weaknesses of a school district's curriculum, and can be used to modify instructional practice. Results have been used for the Michigan Accreditation Program, and will continue to be used as one piece of this program as it evolves into an accountability model.

MEAP vs. Other Tests

Michigan's MEAP tests are based on the Model Core Curriculum Outcomes and the Content Standards approved by the Michigan State Board of Education. No other published tests match Michigan's Outcomes and Standards. Most

MEAP test questions have actually been written by Michigan educators. Also, Michigan's MEAP tests are criterion-referenced, meaning that the results are reported as performance against a standard. These standards are set by Michigan educators and approved by the Michigan State Board of Education. Student performance is judged according to whether or not each student met the achievement standard. If a student meets the standard, it means he/she meets expectations set by the State Board of Education on the recommended curriculum. In theory, all students in the state could achieve the standard in every subject.

Most [other] published tests are, [by contrast], norm-referenced. This means that each student's performance is compared to other students' performance, and *not* to expectations set by educators. No matter how well students do on a norm-referenced test, half of them will always be "below average," even if they meet expectations. For example, imagine a foot race involving 100 people. The person who finishes first performed better than the other 99 participants. Every person who races is ranked-ordered by the time it took them to finish. Someone must finish first, and someone must finish last... but only half of the people can finish in the top 50%.....

Test Development

Test development is a painstaking process. The first step is to have a curriculum upon which the test is based. Our current tests are based on various curricula approved by the State Board of Education.

- The **Essential Skills Reading and Mathematics tests** for 4th and 7th graders are based on the Michigan Essential Goals and Objectives for Reading and Mathematics, approved in 1986 and 1988, respectively. The tests for these subjects were not administered until 1989 and 1991, a full three years after the content was approved. That's about how long planning for a test usually takes.

- The **Science tests** for 5th and 8th graders are based on goals and objectives approved in 1991.

- The **Social Studies** tests are based on Content Standards and Benchmarks approved in 1995.

- The **Writing tests** are based on goals and objectives approved in 1985.

Once a curriculum is approved, the MEAP Assessment Office oversees the development of an Assessment Plan. This plan is carried out in one of several ways. Professional organizations (such as the Michigan Council of Teachers of English, Michigan Council of Teachers of Mathematics, Michigan Reading Association, and the Michigan Science Teachers Association) are sometimes contracted to develop a test "blueprint." At other times, various committees are responsible for this task.

For 2001-2002 test development, assessment planning committees were comprised of educators from across Michigan who drafted a blueprint. The committees sought to determine:

- What is testable?

- How many and what types of questions will be used?

- How will the scoring occur?

- What will score reports look like?

Drafts, comments from the field, and recommended changes are the taken to the State Board of Education, which has the right to approve or change the drafts. Once an approved plan is in place, a test development contract is put out for bid.

Test development contracts call for development, tryouts, and piloting of all new test questions, and include creating scoring guides for all constructed-response items. Bidders on these contracts have historically included major publishers like American College Test (ACT), Advanced Systems, CTB McGraw Hill, Harcourt Educational Measurement, Measurement Incorporated, National Computer Systems, and Riverside Publishing. These companies have large staffs experienced in developing customized assessments for states, school districts, and the nation.
In some cases, the publishers write all of the test items. In other cases, Michigan teachers write the items, and then the publishers proofread, edit, and assemble the pilot test forms. Both options have advantages and disadvantages, and

both have produced good test questions in the past.
After an initial pool of items is written, the items are taken to two different committees for review.

- The **Content Committee** is comprised of educators from across Michigan at the grade levels to be tested. The committee assures that the items are grade-level appropriate, and that they match the content standards. Items that don't match the curriculum are not allowed on the test. The committee often edits items, makes improvements, and discards inappropriate questions.

- The **Bias Review Committee** has the daunting task of reviewing each reading passage, each writing prompt, each science investigation, each test scenario, and each question to assure fairness for all students. The committee may reject items it considers inappropriate for Michigan students.

After reviews are complete and items approved, the items are put in forms for "pilot testing." Schools selected at random are asked to "pre-test" the items. Although individual student results at this stage are not the focus, it is important that students put forth their best effort. Pilot-test data is important in deciding whether an item actually becomes part of an "item bank." Teacher comment sheets, collected during pilot-testing, give the MEAP office another opportunity to receive valuable feedback about the items. Being selected as a "pilot" school gives the opportunity to offer constructive feedback to the MEAP staff through students' performance and through teachers' comments.

Validity of Test Items
The MEAP office looks at data in many ways to assure items are measuring what they are intended to measure. One of the first criterion is whether an item appropriately tests the content. It is difficult for Content Committees to know with certainty that an item adequately addresses content simply by looking at the item. The data from tryouts and pilots offers invaluable insight.

- **p-Value** - The first piece of data examined is called the "p-value." It tells the MEAP office the percentage of students who answered the item correctly. The MEAP staff also looks at the percent of students who chose each "distracter" (incorrect answers on a multiple-choice test). Particular attention is paid when less than 30% of the students select the correct answer. Since all multiple-choice items on MEAP tests have four options, chance alone says that 25% of the students should mark the correct answer. Even if the content is appropriate, the item may not be measuring well... perhaps the graphic shown on the test is somehow misleading, or the question is poorly worded. P-value data is not the final decision on an item... it is simply used to indicate the need for further review.

- **DIF** - Differential Item Functioning is a fancy way of saying an item is potentially biased, or that it functions differently for one group than it does for another. If an item is identified as being potentially biased, it is returned to the Bias Review Committee. Sometimes the content of the item is really a curricular issue, meaning that one group did not do as well as another because they hadn't been taught the material. Perhaps something in the context of the item was missed in the first round of reviews. Again, items are usually allowed to remain, revised, or discarded based on the decisions of the review committees. Changes to an item necessitate that it be pilot-tested again before it may appear on an operational test.

- **Discrimination** - Item discrimination examines performance between students who score high on the test as compared to those who score low. If an item discriminates poorly, it means that low-scoring students did as well or better than high-scoring students. This often occurs on very easy items that practically everyone answers correctly. As long as the item is measuring good content, an item that discriminates poorly is kept. However, if more low-scoring students do as well or better than high-scoring students on a moderately difficult or difficult item, the item is given a closer look by the MEAP staff. Perhaps there is more than one correct answer, or perhaps something in the knowledge-base of the high-scoring students is interfering with the way they are answering the question. The MEAP staff also look at the distracters to assure they are not misleading students in ways unintended.

- **Range** - While variety may be "the spice of life," it is also an important part of testing. The MEAP staff aggressively seek a wide range of difficulty in items. There is, however, no "magic formula" for how many "difficult" or how many "easy" questions are used. A sincere attempt is made to make questions used one year similar to those used the next. The MEAP staff do everything they can to help assure that differences

from one year to the next are small.

- **Other Factors** - For constructed-response items, the staff examines the percent of students receiving points at each score level. If no one is receiving the top score possible, the staff takes another look at what the question is asking. The staff also considers consistency among those who scorer (or "grade") the tests. If an item is not being scored reliably, the staff assesses whether something is wrong with the item, or with the training of those who score the tests.

The selection of items appearing on a test is often done by committee. In Mathematics, for example, the Content Committee examines each objective to be tested in the coming year, then looks at all the items in the item bank that match the objective. They select the item they wish to use to measure the objective. They try not to use the same items over and over, and they try to make sure they don't pick only the most difficult (or easiest) items. The committee also reviews items about which complaints have been received, and either revise or discard the items. Changes to an item necessitate that it be pilot-tested again before it may appear on an operational test.

Rangefinding and Scoring

It is difficult for many to believe that the MEAP constructed-response test items can be scored objectively. But it's even more difficult to understand how this is accomplished without the discussion getting rather technical.
When items are pilot-tested, the MEAP staff uses pilot scoring guides (or "rubrics"). The rubrics undergo revisions. The staff tries to assure fairness and consistency, while giving each student "the benefit of the doubt."
For Mathematics, Social Studies and Science, the scoring guides are item-specific. This means that the scoring guide is written specifically for that item. For Reading and Writing, the scoring guide is consistent from year to year. These scoring guides appear on the back of student test booklets, and may be photocopied and used by school staff and students at any time.

At both the pilot-test and operational stages, the beginning process in scoring is called "rangefinding." During rangefinding, a group of educators (mostly classroom teachers) from across Michigan gather in Lansing to help MEAP staff establish the "range" (or criteria) for open-ended items. The scoring contractor and Michigan Department of Education staff participate in these meetings, but the final decisions are made by the educators.
Participants typically score 160 papers from a sample of student papers. The papers contain real work from real students, from a geographic and ethnic balance around the state. Every single paper is discussed until a consensus is reached on the score the paper should receive. Some papers are easier to score than others, and require little discussion. Some lead to lengthy, spirited discussions because group members are divided in their opinions of what score to give (for example, a two and a three).

In Math, Science, and Social Studies, changes to the scoring rubrics can occur during rangefinding. Sometimes an item does not elicit the kind of response the author intended to receive. When that happens, the scoring guide is adjusted to give students the benefit of the doubt. In pilot rangefinding, problems with items often lead to improvements in the design of the questions.

The papers used in rangefinding are used in a variety of ways:

- To train scorers
- To assess whether the scorers have learned the criteria
- To constantly reevaluate the scorers during scoring

All "handscoring" is currently performed by Measurement Incorporated (known as "MI"), a company based in Durham, North Carolina. MI recruits, hires, and trains Michigan educators to work at scoring sites in Grand Rapids and in Ypsilanti. Before being hired, potential scorers undergo an interview process and write an essay. Once hired to score a particular project, scorers must prove they can match the criteria as established by Michigan educators.
Scorers are trained using rangefinding papers. They generally begin with the scoring guide and clear examples of papers receiving scores of one, two, three, and four. Trainers clearly describe why a certain paper received a certain score. General rules are explained. If there's more than one way to arrive at a correct answer, those options are carefully reviewed. Initial training and practice can take one or two days. Then, scorers begin taking qualifying sets. Qualifying sets are done individually. Each scorer scores a set of papers that were used in rangefinding. Acceptable scores are known only to the trainers. Scorers must be able to match the acceptable score. If not, they receive more

training. If scorers do not qualify, even after retraining, they are dismissed. No student work is ever scored by a scorer who has not passed qualifying tests. Once they have qualified, they begin to score actual work by students.
With few exceptions, papers are scored by two scorers. Scorers do not know the student's name, gender, ethnicity, or hometown. If the first and second scorers agree upon a score, the student receives that score. If, however, the two scorers disagree by one point, the average of the two scores is used. If they disagree by more than one point, the paper is sent to a third scorer (table leader or scoring director) for resolution. Resolutions are rare.
Monitoring by trainers doesn't end just because a scorer has qualified. Other papers from rangefinding are used for validity sets, where scorers are tested while they are scoring. Every day during scoring, the MEAP staff receives faxes which indicate for each scorer how many papers they have scored, how many times they agreed with a second scorer, how many times they were within a point of a second scorer, and whether the score was lower or higher than a second scorer's score. the MEAP staff tracks the percent agreement on a daily basis and over time. This information is used to monitor the scoring process.

Standard-Setting
Standard-setting is the process of determining "cut" scores that allow student performance to be divided into categories.
4th & 7th Grade categories

Satisfactory
Moderate
Low

5th & 8th Grade categories (Science)

Proficient
Novice
Not Yet Novice

5th & 8th Grade categories (Writing)

Proficient
Not Yet Proficient

5th & 8th Grade categories (Social Studies)

Level 1 (Exceeded Michigan standards)
Level 2 (Met Michigan standards)
Level 3 (Basic)
Level 4 (Apprentice)

High School categories

Level 1 (Exceeded Michigan standards)
Level 2 (Met Michigan standards)
Level 3 (Basic)
Level 4 (Unendorsed)

Although the State Board of Education has the final responsibility for approving standards, input to the Board in the form of recommendations typically comes from the standard-setting process.

Standard-setting begins with the selection of a representative committee of educators from across Michigan. Nomination forms are sent to educational organizations, distributed at meetings, and included in the *MEAP Update* (distributed to MEAP Coordinators and all school principals by the MEAP office in Lansing). Most standard-setting panelists are classroom teachers. Others include administrators, curriculum specialists, counselors, parents, and business leaders. Committees represent the geographic and ethnic diversity of our state.

Committees participate in a three-day process which allows them to rate student performance. After three days, the committee's recommendations for "cut" scores are taken to the Bias Review Committee, the Content Advisory Committee, the Assessment Advisory Committee and the Technical Advisory Committee for review. Finally, the cut scores are taken to the State Board of Education for approval.

Initially, the facilitator reviews the charge to the Standard-setting Committee, including how the members (judges) are

selected, and the purpose of the standards. The judges are also briefed on the standard-setting process, and are encouraged to ask questions about the tests and the upcoming tasks. After a description of the assessment development process and an overview of the test content, each judge takes the test. Directions are read to the judges just as they would be to actual students. Judges are given no more than the same tools students would be permitted to use. The judges try to answer every item. After the test is given, the judges score their own tests. As part of the process, the judges are trained on the scoring criteria for the open-ended items through the use of papers illustrating each score point.

Judges are then given student performance descriptions for each score category, which were developed and reviewed by each subject's Content Advisory Committee. The descriptions reflect what students are expected to know and be able to achieve, given the curriculum upon which the tests are based. The standard-setting facilitator then divides the committee into small groups, and allows them to discuss, interpret, and expand upon the descriptions. Each group reports back to the full group, building a common understanding of the descriptions. These performance definitions are extremely important because they take curricular expectations and test questions, and translate them into performance expectations on the test.

Next, judges examine the test items and accompanying student work. Questions are ordered from the "easiest" to the most "difficult." The task of the judge is to decide, for example, the point at which students get the item right (meet the standard), but will get the next item wrong. This point is then translated into a "cut" score for that judge. Each judge chooses cut scores for all categories. The average becomes the cut score for the entire group. Judges participate in several rounds of ratings. Included are group discussions (both small and large), and the receipt of relevant data. The judges know what percent of students across the state got each item correct, and how many points students received on the open-ended items. They also are given "impact" data, which shows the impact of the cut scores established for the entire state, and on each gender and ethnic group.

Judges know group recommendations at every round. Finally, after all information has been reviewed and discussed, judges make their final ratings. The averages of final ratings are taken forward as the group's recommendations. As part of this process, the MEAP staff monitors both intra- and inter-judge consistency. The standard deviations of the judges almost always drop from one round to the next. Most judges are extremely consistent. Their ratings vary only slightly from round to round. Committee members are also asked to evaluate the standard-setting process. Those who have participated previously are overwhelmingly positive about the experience. They understand that, other than providing administration, the MEAP staff does not participate in the process, nor does it try to influence judges to take a particular point of view. From test-planning, to item development, to scoring, to standard-setting, educators from around the state are involved.

Reliability

Two important technical concepts in measurement are reliability and validity. ...For the MEAP tests, reliability values are determined by using internal consistency formulas, which indicate how homogeneous items are in a test, or the degree to which students' responses to each item correlate with their total test scores. Cronbach's Coefficient Alpha is a measure of internal consistency reliability usually used when constructed response items appear on a test. It can also be used when there are solely multiple-choice items, or when combinations of item types are used. Typically, the more lengthy the test, the higher the reliability.

Below are the reliability indices for the MEAP tests given in 1998-99.

MEAP Reliability 1998-99

Test	Reliability
Grade 4 Reading – Story	.814
Grade 4 Reading - Informational	.809
Grade 4 Mathematics	.931
Grade 5 Science	.886
Grade 5 Social Studies	

	.882
Grade 7 Reading - Story	.891
Grade 7 Reading - Informational	.902
Grade 7 Mathematics	.962
Grade 8 Science	.892
Grade 8 Social Studies	.883
Grade 11 Reading	.830
Grade 11 Mathematics	.892
Grade 11 Science	.878
Grade 11 Social Studies	.888
Grade 11 Writing	.610

Note - Reliability cannot be calculated for a one-item test, so none is provided for 5th and 8th Grade Writing.

Scorer Agreement
All constructed response ("open-ended") answers are scored by two scorers. If the two scorers disagree by more than one point, a third scorer is used. The third scoring is called a "resolution reading." The less frequent the need for resolution, the more accurate the scoring consistency. [For the 5th and 8th MEAP tests in 1998-1999, a third reading was needed less than two percent of the time for about 85 percent of the items.]

Both the reliability and the agreement rates are technically sound for the tests.

Validity
Validity answers the question of whether a test measures what it is supposed to measure. It refers to the degree of appropriateness, meaningfulness, and usefulness of the specific inferences made from test scores.
There are three kinds of validity discussed in Standards for Educational and Psychological Testing (AERA-APA-NCME, 1985, updated 1999):

- Content validity
- Criterion validity
- Construct validity

The current generation of MEAP assessments are based on the Michigan Essential Goals and Objectives for Mathematics Education, Reading Education, Science Education and Writing Education, which were approved by the State Board of Education in 1988, 1986, 1991, and 1985, respectively. The Social Studies test is based on the Michigan Curriculum Framework.

Because the current MEAP assessments are achievement tests used to assess what students have learned and should be able to achieve in specific content areas by the end of a certain grade, the most important type of validity of concern is Content validity. To verify Content validity, test items must match the specified objectives given in the test blueprint or assessment framework.

Like all published achievement tests, the MEAP assessments have a blueprint that indicates the objectives to be tested in each content area. There is an infinite number of ways to write test items to measure each objective, and multiple forms are composed for each test. Not all objectives are tested in any given form of a test. Both "easy" and "difficult"

items are used in every form to balance the difficulty level of the items, and to equate the different forms to one another. The sample of items chosen for a test represents the domain of all possible test items that fit the blueprint. For a student to do well on a test, he/she must have mastered the entire domain, not simply bits and pieces.
Content Advisory Committees, which include teachers and curriculum coordinators, verify that each test question meets the objective it is supposed to measure, and that it fits the blueprint or framework. A Bias Review Committee then verifies that the items are not disadvantaging any particular group. The groups ensure that the MEAP tests have Content validity.

Two other types of validity that psychometricians are often concerned about are Criterion and Construct validity. Criterion validity refers to whether a measure can predict a student's future performance. For example, the ACT and the SAT are used to predict college success, thus Criterion validity is very important for those tests. The publishers of the ACT and the SAT conduct studies to correlate the scores with college grades to ensure they are valid. This is not, however, the purpose of the MEAP High School Test (HST).

Instead, the purpose of the MEAP HST is to determine whether a student is eligible to earn a transcript endorsement in a specific content area. To establish Criterion validity, the MEAP HST would have to be correlated with some other existing measure of student performance. Unfortunately, tests such as the ACT, Advance Placement exams, and the National Assessment of Educational Progress (NAEP) are not based on the same subject matter as the MEAP HST. In some cases, Michigan's curriculum is far more demanding. In other cases, these national tests are either more specific or not appropriate to a student who is not college-bound.

Construct validity is concerned with the parts (or dimensions) of a test, and whether they relate to the construct under study in a MEAP test. The Mathematics test has several strands; the Science and Social Studies tests have multiple dimensions; and the Reading test also has more than one component. A Construct validity analysis (such as a factor analysis, or a structural equation model) could show whether questions fit into particular strands. For example, a Construct validity analysis could answer the question of whether all geometry items on a test are most strongly related to one another, or if one item better fits with data analysis questions.

The results of all MEAP tests, and all decisions made from these results, are based on the total test score, not on scores of an individual strand or dimension. The Rasch model in Item Response Theory is used to equate and scale all MEAP tests. Item Response Theory assumes that the tests under study are "unidimensional." This means that the tests measure one construct (or one domain) only, such as mathematics. Other types of Construct validity analysis could show whether the strand or dimension structure holds, as suggested by the goals and objectives.

The dilemma of whether to estimate Construct validity on the basis of the total score, or upon strand scores, is one with which psychometricians (including those in the MEAP office) constantly struggle. The MEAP office contracts and consults with a Technical Advisory Committee comprised of nationally-known psychometricians who offer advice on such issues. The MEAP staff has always followed, and will continue to follow, current psychometric practice in developing, administering, analyzing, and scoring the Michigan Educational Assessment Program tests.

Endnotes

[1] For information on U.S. lags in student achievement see: Beaton, Albert et al. (1996) Mathematics Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study. CSTEEP, Boston College., Boston MA., 1-186. or John Bishop, "Secondary Education in the United States: What can others learn from our mistakes?" (with Ferran Mane) presented at Challenges of Secondary Education in Latin America and the Caribbean, Inter-American Development Bank's Regional Policy Dialogue, Washington, DC, Dec. 5-6, 2000.

[2] MCEs typically set a low minimum standard. In 1996 only 4 of the 17 states with MCEs targeted their graduation exams at a 10th grade proficiency level or higher. Failure rates for students taking the test for the first time varied a great deal: from a high of 46% in Texas, 34 % in Virginia, 30% in Tennessee and 27% in New Jersey to a low of 7% for Mississippi. However, since students can take the tests multiple times, eventual pass rates for the Class of 1995 were much higher: 98% in Louisiana, Maryland, New York, North Carolina and Ohio; 96 % in Nevada and New Jersey, 91% in Texas and 83% in Georgia. [American Federation of Teachers, *Making Standards Matter:1996* (Washington, DC: American Federation of Teachers, 1996) p. 30.] Since the tests are designed to determine who falls below a pretty low standard, they typically do not assess material that college bound students study in 10th and 11th grade (e.g. Algebra II and geometry proofs). This may result in too much class time being devoted to practicing low level skills.

[3] Testimony of Mark Murray, State Treasurer of Michigan, before the Subcommittee on Oversight and investigations, Committee on Education and Workforce, US House of Representatives, September 8, 1999. Available at www.meritaward.state.mi.us/merit/murrray.htm.

[4] Michigan spent $12,698,697,000 on K-12 education in fiscal 1996. The $100,000,000 annual cost of the Merit Award program is less than 0.8 percent of this figure.

[5] Laurence Steinberg, Bradford Brown, and Sanford Dornbusch, *Beyond the Classroom* (New York: Simon and Schuster, 1996), pp. 145-146.

[6] James A. Kulik and Chen-Lin Kulik, "Effects of Accelerated Instruction on Students," *Review of Educational Research*, Vol. 54 No. 3 (Fall 1984), pp. 409-425; David Monk, "Subject Area Preparation of Secondary Mathematics and Science Teachers and Student Achievement," *Economics of Education Review*, Vol. 13 No. 2 (1994), pp. 125-145. and John H. Bishop, "Incentives to study and the organization of secondary instruction." *Assessing Educational Practices*, eds. William Baumol and Becker (Cambridge, Mass.: MIT Press, 1996), pp. 99-160.

[7] Longitudinal Survey of American Youth, "Data File User's Manual" Q. AA37N.

[8] Interview with counselor at a wealthy suburban school, August 1997

[9] Peter D. Hart Research Associates, "Valuable Views: A public opinion research report on the views of AFT teachers on professional issues" (Washington D.C.: American Federation of Teachers, 1995), pp. 1-24.

[10] National Center for Education Statistics, Digest of Education Statistics, 1993, p. 137 and 1998 p. 329.

[11] James Rosenbaum, "College for all: Do Students understand what College demands?" Northwestern University, (1999), 1-36. and Clifford Adelman, "Answers in the Tool Box: Academic Intensity, Attendance Patterns and Bachelors Degree Attainment." Wash, DC: Office of Educational Research and Improvement, 1999, 1-87.

[12] These tabulations do not measure the causal effects of either schooling or prose literacy. Causal effects will be smaller because early literacy levels influence completed schooling, because additional schooling raises literacy and because working in white collar and professional and managerial jobs raises literacy and increases the probability of returning to school for further education.

[13] Large as it is, this 16 percent figure substantially understates the total effect of improved K-12 learning on earnings as an adult. First, test scores influence hours of work and the risk of unemployment, not just wage rates. Secondly, the AFQT is an incomplete measure of what students are learning in high school. If reliable measures of other skills learned in school (such as science, social studies, writing, technical and computer skills) were included in the model, the total effect of test scores would be larger. The third and most important source of bias comes from using a contemporaneous measure of schooling as a control. Much of the benefit of learning in the first 12 years of school comes from the assistance it provides in continuing schooling beyond high school. Yet, this benefit of learning in high school does not get picked up by the AFQT coefficient. It is captured, instead, by the coefficient on the contemporaneous measure of schooling. If a prospective measure of schooling (completed schooling at the time of the AFQT test) were substituted for the contemporaneous measure, the coefficient on the AFQT would have been much larger. Joseph Altonji and Charles Pierret, "Employer Learning and Statistical Discrimination." Quarterly Journal of Economics

[14] M. H. Brenner. "The use of high school data to predict work performance," *The Journal of Applied Psychology* Vol. 52, # 1, (1968), pp. 29-30.; Department of Labor, *General Aptitude Test Battery Manual* (Superintendent of Documents, 1970).; John E. Hunter, James J. Crosson and David H. Friedman, "The Validity of the Armed Services Vocational Aptitude Battery (ASVAB) For Civilian and Military Job Performance" (Department of Defense, August 1985). John Hartigan and Alexandra Wigdor, eds. *Fairness in Employment Testing* (Washington, D.C.: National Academy Press, 1989). John H. Bishop, "Impact of Academic Competencies on Wages, Unemployment and Job Performance," *Carnegie-Rochester Conference Series on Public Policy*, Volume 37, (December 1992), pp. 127-194.

[15] J. C. Hauser and Thomas M. Daymont, "Schooling, ability and earnings: Cross-sectional evidence 8-14 years after high school," *Sociology of Education*, Vol. 50 (July 1977), 182-206; Paul Taubman and Terence Wales, "Education as an investment and a screening device," *Education, Income and Human Behavior*, ed. F. T. Juster, (New York: McGraw Hill, 1975), pp. 95-122; and Henry Farber and Robert Gibbons, "Learning and Wage Dynamics," *Quarterly Journal of Economics* (1996), pp. 1007-47.

[16] "I Have a Dream" programs also have other important program elements not currently a part of the Merit Award program.. Mehan, H., Hubbard, L. and Villanueva, I. (1994) "Forming academic identities: Accomodation without assimilation among involuntary minorities." Anthropology and Education Quarterly , 25(2), 91-117. and Joseph Kahane and Kim Bailey, "The Role of Social Capital in Youth Development: The Case of 'I Have a Dream' Programs," Educational Evaluation and Policy Analysis , 21(3), Fall 1999, 321-343.

[17] In the educational context the phrase "high stakes" refers to decisions that have big effects on a person's life. Examples of such decisions are classification as needing special education, retention in grade, the award or denial of a high school diploma and admission to state colleges. One way to measure the stakes is to calculate impacts on lifetime earnings of completing extra years of schooling. The present

discounted value of these earnings differentials is roughly $200,000 for high school graduation and $525,000 for college graduation. These are decisions that Michigan educational institutions base on multiple indicators of student achievement and aptitude of which high school grades are the most important..

[18] The stakes attached to getting a merit award are smaller than those associated with getting a 4 or 5 on an Advanced Placement (AP) exam. Students who get a 4 or 5 on an AP exam are typically awarded 3 to 8 credits by the college they attend. Note that the grades awarded by teachers have no impact on whether one gets advanced placement credit from a postsecondary institution. Everything depends on how the student does on a single 3 hour exam. At Cornell University tuition is $750 per credit, so students potentially save $3000 when awarded 4 AP credits and $6000 when awarded 8 AP credits. At low tuition public colleges the tuition savings are smaller, but the saving in time is just as important. If one gets 16 AP credits from one's college, one can graduate and enter the labor market as a college graduate one semester earlier. During that 5 month period one can earn $10,000 to $15,000. If the Michigan Merit Award Program is forced to use teacher grades as one of the criteria in its awards because the stakes are considered to be high can a challenge to public colleges awarding advanced placement credit based on AP exams be far behind?

[19] Michigan had 546,000 students in colleges and universities in fall 1996. Assuming that two-thirds of these students claim an average credit of $1000, I estimate that about $366,000,000 in federal education tax credits were awarded to Michigan families in fiscal 2000.

[20] Barbett , Samuel and Korb, Roslyn. Current Fund Revenues and Expenditures of Degree Granting Institutions: Fiscal Year 1996 , Washington, DC: National Center for Education Statistics, NCES 1999-161, 1-40.

[21] This observation is based on interviews with the directors of the testing and accountability divisions in Manitoba and New Brunswick Canada and the large increases in student performance that occurred in New Brunswick, Massachusetts, Michigan and other states when no-stakes tests become moderate or high-stakes tests (Ed Hayward, "Dramatic Improvement in MCAS scores" Boston Herald, Oct. 16, 2001). Experimental studies confirm the observation. In Candace Brooks-Cooper master's thesis, a test containing complex and cognitively demanding items from the NAEP history and literature tests and the adult literacy test was given to high school students recruited to stay after school by the promise of a $10.00 payment for taking a test. Students were randomly assigned to rooms and one group was promised a payment of $1.00 for every correct answer greater than 65 percent correct. This group did significantly better than the students in the other test taking conditions, one of which was the standard try your best condition. Candace Brooks-Cooper, 1998. Similar results were obtained in other well designed studies conducted by the National Center for Research on Evaluation, Standards and Student Testing: Vonda Kiplinger and Robert Linn, "Raising the Stakes of Test Administration: The Impact on Student Performance on NAEP," CSE Technical Report 360, March 1993, 1-72. and Harold F. O'Neil, Renda Sugre, Jamal Abedi, Eva L. Baker, and Shari Golan, "Final Report of Experimental Studies on Motivation and NAEP Test Performance," CSE Technical Report 427, June 1997, 1-176.

[22] The SAT-I and the ACT fail to assess most of the material--economics, civics, literature, foreign languages and the ability to write an essay--that high school students are expected to learn. The SAT-I leaves history and science out as well. The ACT's science and history subtests are very short and not linked to specific curricula. They are as much a reading test as a test of content knowledge in science and history.

[23] The models controlled for East Asian nation and for GDP per capita. John H. Bishop, (1996) "The Impact of Curriculum-Based External Examinations on School Priorities and Student Learning." International Journal of Education Research; John H. Bishop, "The Effect of National Standards and Curriculum-Based External Exams on Student Achievement." American Economic Review, May 1997, Similar results were obtained by Ludger Wößmann, "Schooling Resources, Educational Institutions, and Student Performance: The International Evidence," Kiel Working Paper No. 983, (May 2000) Kiel Institute of World Economics, Germany, <http://www.uni-kiel.de/ifw/pub/kap/2000/kap983.htm> 1-88.

[24] John H. Bishop, "Are National Exit Examinations Important For Educational Efficiency?" Swedish Economic Policy Review, Vol. 6, #2, Fall 1999, 349-401.

[25] Bishop John. "Do Curriculum-Based External Exit Exam Systems Enhance Student Achievement?" University of Pennsylvania, Consortium for Policy research in Education, CPRE Research Report RR-40, 1998, 1-32. John H. Bishop, "Nerd Harrassment, Incentives, School Priorities and Learning," *Earning and Learning*, ed. by Susan Mayer and Paul Peterson, (Washington, DC: Brookings Institution Press, 1999a). John H. Bishop, "Are National Exit Examinations Important For Educational Efficiency?" Swedish Economic Policy Review, Vol. 6, #2, Fall 1999, 349-401.

[26] Ibid.

[27] John H. Bishop, "Nerd Harrassment and Grade Inflation: Are College Admissions Policies Partly Responsible?" Center for Advanced Human Resources Discussion Paper #99-14, (1999c).

[28] Grissmer et al (2000) carefully adjusted state NAEP trend data for changes in the ethnic and socio-economic composition of the students taking NAEP assessments in the state and calculated corrected estimates of the annual rate of test score gains in the state. Exclusion of students from testing was also analyzed and adjusted for. States had to have participated in successive state NAEP assessments in the same subject area to be included in the analysis. There 36 states that met this criterion. Other states with particularly large test score gains were: Texas (# 2), Michigan (# 3), Indiana (# 4), Maryland (# 5) and West Virginia (# 6). Since there were no stakes for students in Michigan during this period, other factors such as the school accountability system or increased spending probably account for the success.

[29] The letter that invited schools to participate in the study and join the Educational Excellence Alliance was worded as follows: "We are writing to offer your school the opportunity to obtain an assessment of student norms and peer culture at absolutely no cost to the school. The assessment is being undertaken by the Educational Excellence Alliance (EEA), a group of striving high schools that are interested in learning how to help all their students to achieve at higher levels. To join EEA all you need to do is to administer the enclosed questionnaire to your tenth graders and complete a short questionnaire about the school. We will scan the questionnaire, tabulate the answers and report back to you how the tenth grade answered each question and how their culture and norms compare with that of other schools serving students with similar socio-economic backgrounds. This report should allow you and your staff to more intelligently plan your efforts to improve achievement and build a student culture that honors academic achievement and respects individual differences...." This letter was sent to all high school principals and superintendents of schools in Connecticut and the superintendents and principals of high schools in northern New Jersey (Essex, Bergen, Hudson, Passaic and Morris counties), in Berkshire, Essex, Hampden, Norfolk, Middlesex,, Plymouth and Worcester counties in Massachusetts, Albany, Broome, Duchess, Erie, Nassau, Niagara, Oneida, Onondaga, Orange, Oswego, Putnam, Rensselaer, Rockland, Saratoga, Schenectady, Suffolk, Westchester counties in New York. New York City was approached but chose not to participate. In addition invitations were sent to all non-public high schools in New York State outside of New York City and in the seven targeted Massachusetts counties. Ten to fifteen percent of invited schools

agreed to participate in the study and returned their questionnaires in time to be included in this analysis.

[30] The states neighboring New York test students during high school but the tests are not given as part of a course and scores on the tests are not part of the student's grade. New Jersey's tests are first given in October of 11th grade and passing scores in reading, writing and math are required to graduate. The Connecticut and Massachusetts tests are first administered in May of 10th grade. Connecticut puts scores on transcripts and awards Certificates of Mastery to the 40 percent of students who exceed state goals. The Massachusetts tests become a graduation requirement for students entering high school in Fall 1999. The Massachusetts 10th graders surveyed for this study during the 1998/99 school year were scheduled to take the test under no fault conditions.

[31] California is not counted as a CBEEES state because (a) the state did not have a MCE graduation requirement, (b) teachers could not use Golden State exam scores in their own grading, (c) other rewards for doing well on the exams were weak and (d) the program was being phased in slowly so by the middle of the 1990s most students were not participating and most participating teachers had not been teaching in the new environment long enough to change their expectations of what students were to achieve.

[32] Michigan's Merit Award program is pretty unique. Ohio is the only state with a similarly structured merit award program. It is a very new program and it's awards are only $500. Another difference is that, unlike Michigan, Ohio has a minimum competency exam graduation requirement. Consequently, it will be difficult to draw lessons from Ohio's experience.

[33] The ratio of HST test takers to 12th graders expected to graduate in 2000 is below 1.0 for four reasons. Some students are exempted at parent request or because of special education or LEP status or are absent on the day of the testing. Another reason might be 12th graders who do not have enough Carnegie units to graduate with their class.

[34] Susan Dynarski, David Mustard ….

[35] I use data on the sum of undergraduate and graduate enrollment because separate data on undergraduates or freshman are not yet available for 2000 and 2001. Data on total enrollment including non-resident enrollment was used because some universities changed their definition of 'resident' student during this period.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: A Prospective Policy Evaluation of the Michigan Merit Award Program	EA 032 097
Author(s): John H. Bishop	
Corporate Source:	Publication Date: June 2002

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

Level 1

☒

Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.

The sample sticker shown below will be affixed to all Level 2A documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2A

Level 2A

☐

Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only.

The sample sticker shown below will be affixed to all Level 2B documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2B

Level 2B

☐

Check here for Level 2B release, permitting reproduction and dissemination in microfiche only.

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, please

Signature: John H Bishop	Printed Name/Position/Title: John H Bishop
Organization/Address:	Telephone: 607-255-2742 FAX: 607-255-1836
	E-Mail Address: jhb5@cornell.edu Date: 5/31/03

(Over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, *or*, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

ERIC Clearinghouse on Educational Management
5207 University of Oregon
Eugene OR 97403-5207
Attn: Acquisitions

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
4483-A Forbes Boulevard
Lanham, Maryland 20706

Telephone: 301-552-4200
Toll Free: 800-799-3742
FAX: 301-552-4700
e-mail: ericfac@inet.ed.gov
WWW: http://ericfacility.org

EFF-088 (Rev. 2/2001)